Genomic Targets and Features of BarA-UvrY (-SirA) Signal Transduction Systems

Tesfalem R. Zere[1¤a], Christopher A. Vakulskas[1¤b], Yuanyuan Leng[1], Archana Pannuri[1], Anastasia H. Potts[1], Raquel Dias[1], Dongjie Tang[1¤c], Bryan Kolaczkowski[1], Dimitris Georgellis[2], Brian M. M. Ahmer[3], Tony Romeo[1] *

1 Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, United States of America, 2 Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México D.F., México, 3 Department of Microbial Infection and Immunity, The Ohio State University, Columbus, OH, United States of America

¤a Current address: Department of Oral Biology, University of Florida, College of Dentistry, Gainesville, FL, 32610–0424
¤b Current address: Integrated DNA Technologies, Molecular Genetics Department, 1710 Commercial Park, Coralville, IA, 52241
¤c Current address: State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, and College of Life Science and Technology, Guangxi University, Nanning, Guangxi, PR China
* tromeo@ufl.edu

Abstract

The two-component signal transduction system BarA-UvrY of *Escherichia coli* and its ortho-logs globally regulate metabolism, motility, biofilm formation, stress resistance, virulence of pathogens and quorum sensing by activating the transcription of genes for regulatory sRNAs, e.g. CsrB and CsrC in *E. coli*. These sRNAs act by sequestering the RNA binding protein CsrA (RsmA) away from lower affinity mRNA targets. In this study, we used ChIP-exo to identify, at single nucleotide resolution, genomic sites for UvrY (SirA) binding in *E. coli* and *Salmonella enterica*. The *csrB* and *csrC* genes were the strongest targets of cross-linking, which required UvrY phosphorylation by the BarA sensor kinase. Crosslinking occurred at two sites, an inverted repeat sequence far upstream of the promoter and a site near the -35 sequence. DNAse I footprinting revealed specific binding of UvrY *in vitro* only to the upstream site, indicative of additional binding requirements and/or indirect binding to the downstream site. Additional genes, including *cspA*, encoding the cold-shock RNA-binding protein CspA, showed weaker crosslinking and modest or negligible regulation by UvrY. We conclude that the global effects of UvrY/SirA on gene expression are primarily mediated by activating *csrB* and *csrC* transcription. We also used *in vivo* crosslinking and other experimental approaches to reveal new features of *csrB*/*csrC* regulation by the DeaD and SrmB RNA helicases, IHF, ppGpp and DksA. Finally, the phylogenetic distribution of BarA-UvrY was analyzed and found to be uniquely characteristic of γ-Proteobacteria and strongly anti-correlated with *fliW*, which encodes a protein that binds to CsrA and antagonizes its activity in *Bacillus subtilis*. We propose that BarA-UvrY and orthologous TCS transcribe sRNA antagonists of CsrA throughout the γ-Proteobacteria, but rarely or never perform this function in other species.

Introduction

The ability of bacteria to flourish under diverse environmental conditions requires their physiology and metabolism to be regulated by complex transcriptional and posttranscriptional circuitries. The Csr (carbon storage regulator) or Rsm (repressor of stationary phase metabolites) system is a post-transcriptional regulatory system of *E. coli* and other γ-proteobacteria that is extensively interconnected with transcriptional regulatory circuits [1–8]. Its centerpiece, CsrA, is a small dimeric RNA binding protein that regulates mRNA translation, turnover and transcription termination [3, 9–11]. CsrA activity in *E. coli* and *Salmonella enterica* serovar Typhimurium (hereafter referred to as *Salmonella*) is regulated by CsrB and CsrC sRNAs, which utilize multiple CsrA binding sites to sequester CsrA away from its lower affinity mRNA targets [3, 12–14]. CsrB/C levels are regulated by factors affecting both their synthesis and turnover [15].

CsrB and CsrC transcription is directly regulated by the BarA-UvrY or BarA-SirA two-component signal transduction system (TCS) in *E. coli* [13, 16, 17] and *Salmonella* [14, 18, 19], respectively. The BarA protein belongs to a family of membrane-associated tripartite sensor-kinases and UvrY belongs to the FixJ family of response regulators [14, 19–21]. BarA is required for sensing the presence of acetate, formate and other carboxylate compounds by an undetermined mechanism [22–24]. This leads to autophosphorylation at a conserved histidine residue (His^{302}), and transphosphorylation of UvrY/SirA through a His^{302}→ Asp^{718}→ His^{861}→ Asp^{54} phosphorelay [20, 22]. Phosphorylated UvrY-P/SirA-P, in turn, binds to its DNA targets and regulates their transcription. For instance, *in vitro* binding studies suggest that the SirA protein (SirA-P) binds to DNA sequences located upstream of target genes for Csr sRNAs and activates their transcription in *Salmonella* [18, 19, 25].

BarA-UvrY and its orthologs in other γ-proteobacteria, including GacS/GacA (*Pseudomonas*), VarS/VarA (*Vibrio*), ExpS/ExpA (*Pectobacterium)* and LetS/LetA (*Legionella pneumophila*) have been reported to regulate virulence, metabolism, biofilm formation, stress resistance, quorum sensing and secretion systems [4, 20, 26–30]. Transcriptomics studies of UvrY and its orthologs have shown effects on the expression of numerous genes [30–32]. An understanding of which of these effects represents direct regulation vs. indirect regulation via effects on the Csr sRNAs is necessary for modeling of the complex genetic circuitry that underpins the systems biology of these species. This question has been examined in only one case, which used ChIP-on-chip analysis to conclude that only the genes for Csr sRNAs (RsmY, RsmZ) in *Pseudomonas aeruginosa* are direct targets for GacA binding [33].

Here, we used ChIP-exo, an advanced procedure for determining genome-wide DNA-protein interactions with single nucleotide resolution [34] to map the UvrY and SirA DNA binding sites across the *E. coli* and *Salmonella* genomes. The *csrB* and *csrC* genes were by far the strongest direct targets of binding in both species. These genes exhibited crosslinking at two distinct locations. By comparing the results of *in vivo* crosslinking with *in vitro* DNA binding assays, an 18 NT palindrome or inverted repeat sequence (IR) located far upstream of the promoter was found to serve as a specific binding site for UvrY at *csrB* and *csrC*. Disruption of *barA* eliminated UvrY binding *in vivo*. IHF (integration host factor) was required for optimal UvrY binding to and transcriptional activation of *csrB* but not *csrC*. In addition to the *csrB/C* genes, weaker genomic binding sites of UvrY were identified by ChIP-exo. Several genes associated with the weaker binding sites were tested and found to exhibit negligible or only modest regulation by UvrY, suggesting that diverse effects of this TCS are mediated through Csr circuitry.

In addition to BarA-UvrY, other factors activate *csrB/C* transcription in *E. coli*, such as CsrA [13, 16, 35], DksA and ppGpp of the stringent response [2], and the DEAD-box RNA helicases DeaD and SrmB [4]. The stringent response describes a regulatory network of eubacteria that

responds to amino acid starvation and other stresses [2]. Activation of this system is characterized by a rapid downshift in synthesis of stable RNAs, such as rRNA and tRNA and stimulation of the expression of genes involved in amino acid biosynthesis and transport, although these processes represent a fraction of its global regulatory role [36]. The effector of this response is the nucleotide secondary messenger guanosine tetraphosphate, ppGpp, also known as "magic spot" [37]. Complexed with RNA polymerase, ppGpp positively or negatively affects transcription [38]. In most cases, regulation by ppGpp requires the RNA-polymerase associated transcription factor DksA [39, 40], which acts together with ppGpp to synergistically regulate expression of a number of global regulators and other genes. The DEAD-box RNA helicases are enzymes that utilize ATP energy to alter RNA structure or RNA-protein interactions [4]. Their roles in the posttranscriptional regulation of bacterial gene expression are presently underappreciated. DeaD regulates *uvrY* translation by altering mRNA structure, while SrmB uses a distinct, but incompletely understood mechanism. We used a combination of *in vivo* and *in vitro* approaches to refine our understanding of how these regulatory factors affect *csrB/C* transcription.

Bioinformatics analyses revealed that BarA-UvrY orthologs are strongly anti-correlated with the *fliW* gene, which encodes a protein that binds to and antagonizes CsrA of *B. subtilis* [41], indicating that few if any species use both FliW and BarA-UvrY transcribed sRNAs as CsrA antagonists. These studies advance our understanding of this global regulatory circuitry and highlight the importance of including post-transcriptional regulation along with transcriptional control when modeling bacterial regulatory networks.

Experimental Procedures

Bacterial strains, plasmids and primers used in this project are listed in S1 Table.

Bacterial strains and culture conditions

LB medium (1% [w/v] Tryptone; 1% [w/v] NaCl and 0.5% [w/v] yeast extract) was used for culture of bacteria unless stated otherwise. The antibiotics ampicillin (100 μg mL−1), tetracycline (15 μg mL−1), kanamycin (50 μg mL−1), and chloramphenicol (25 μg mL−1) were included in growth media as needed. Bacterial cultures were stored at -80°C in medium containing ~15% glycerol. To revive cultures, LB medium (2 mL) was inoculated from the frozen stock cultures and incubated with shaking (250 rpm) at 37°C overnight. The overnight cultures served as inoculum (1:1000) for LB medium or Kornberg medium (1.1% [wt/vol] K_2HPO_4, 0.85% [wt/vol] KH_2PO_4, 0.6% [wt/vol] yeast extract containing 0.5% [wt/vol] glucose). Growth was determined by monitoring OD_{600} and/or by assaying total cellular protein.

Construction of chromosomal deletions and FLAG® fusion proteins

Gene deletions (of *E. coli*) and carboxy-terminally 3XFLAG®-tagged constructs (of *E. coli* and *Salmonella*) were introduced by the Red recombinase method, as described [42, 43]. For gene deletions, primers carrying sequences (40 nt) for the beginning (forward primer) or end (reverse primer) of the open reading frame were designed (S1 Table) and plasmids pKD13 (or pKD3) were used as template DNAs for PCR amplification [43]. For constructing 3XFLAG tag constructs, primers carrying sequences (40 nt) matching the terminus of the targeted gene (forward primer) and the region downstream from it (reverse primer) were designed (S1 Table) and plasmid pSUB11 was used as template DNA for PCR amplification [42, 43]. PCR products were gel purified and used directly for electro-transformation. *E. coli* and *Salmonella* strains carrying pKD46 helper plasmid were grown at 30°C in SOB (0.5% [w/v] yeast extract; 2% [w/v] tryptone; 10 mM NaCl, 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$) supplemented with

100 μg/mL ampicillin and 10 mM arabinose to mid-log (OD_{600} of 0.5), collected by centrifugation, washed three times with ice-cold 10% glycerol and resuspended with ice-cold 10% glycerol. The transformation was then performed by electroporation. After 1h recovery at 37°C in SOB medium, bacteria were spread onto LB agar plates supplemented with antibiotics for the selection of Cm^R or Kn^R recombinants. Correct insertion of the marker and 3XFLAG sequence in the genome was confirmed by PCR amplification and by sequencing of the insertion site using primers listed in (S1 Table). When necessary, the FRT-flanked antibiotic resistance cassette was removed using pCP20 [43].

Construction and purification of carboxy-terminally His-tagged UvrY protein

His-tagged UvrY (UvrY-His$_6$) protein was constructed by PCR amplification of the coding sequence of *uvrY* gene using genomic DNA of *E. coli* MG1655 as template DNA and oligonucleotides UvrY-6xhis-F and UvrY-6xhis-R as primers (S1 Table). The amplicon was gel-purified, digested using *Nde* I and *Xho* I restriction enzymes (NEB), and cloned into a similarly digested and dephosphorylated pET24-a (+) vector DNA (NEB) using electrocompetent DH5α for the transformation. The pET-UvrY plasmid was confirmed by gel electrophoresis and sequencing of the inserted DNA using T7-promoter and T7 terminator primers. The pET-UvrY plasmid was moved into BL21 (DE3) *E. coli* strain for expression. The *E. coli* BL21_DE3/pET-UvrY cells carrying UvrY-His$_6$ were grown in 500 mL of LB containing 50 μg/ mL kanamycin at 37°C. At OD_{600} of 0.6, expression of the UvrY-His$_6$ was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 1 mM, followed by 2h of shaking incubation at 37°C, 250 rpm.

The purification procedure for UvrY-His$_6$ was similar to the method previously described for isolation of DeaD-His$_6$ with some modification [4]. Cells were collected by centrifugation (600 x G, 4°C, 5min), suspended in buffer (20 mM Tris-HCl [pH 7.9], 500 mM NaCl, 20 mM imidazole) containing a protease inhibitor cocktail (cOmplete Mini, EDTA-free protease inhibitor cocktail, Roche Diagnostics), and lysed using a French Press. The lysates were centrifuged (20,000 x G, 15 min, 4°C) to remove unbroken cells and cell debris and the supernatant solutions were purified by affinity chromatography on a HisTrap HP column as instructed by the manufacturer (GE Healthcare Life Sciences). Fractions containing UvrY were pooled and dialyzed against UvrY storage buffer (20 mM Tris-HCl (pH 7.9), 150 mM NaCl, 1 mM dithiothreitol and 10% glycerol). UvrY was then aliquoted, frozen using liquid nitrogen, and stored at −80°C. The bicinchoninic acid method was used for assaying protein concentration, as recommended (Pierce Biotechnology).

UvrY phosphorylation *in vitro*

For *in vitro* phosphorylation, the purified UvrY-His$_6$ protein was incubated with 100 mM acetyl-phosphate (Sigma) in a phosphorylation buffer containing 50 mM HEPES pH7.5, 100 mM NaCl and 10mM $MgCl_2$. The reaction mixtures were incubated for 60 min at room temperature. To determine relative amount of UvrY-P formed, the reactions were fractionated on Phos-tag^{TM} SDS-PAGE gels [1.0mm Protean 3 (Bio-Rad) minigels] of the following composition: 7.5% separating gel [7.5% (29:1) acrylamide: bis-acrylamide, 357mM Bis-Tris, pH 6.8, 100 μM Zn $(NO_3)_2$, and 50 μM Phos-tag reagent (Waco Pure Chemical Industries)] with a 4% stacking gel [4% (29:1) acrylamide:bis-acrylamide, 357 mM Bis-Tris, pH 6.8]. UvrY-P was resolved by electrophoresis at constant 150 V at 4°C for 70 min using modified MOPS running buffer (100 mM Tris, 100 mM MOPS, 0.5% SDS, and 5 mM $NaHSO_3$). Gels were subsequently

stained with Coomassie blue and the signals were imaged using a ChemiDoc XRS+ system (Bio-Rad) and quantified using Quantity One image analysis software (Bio-Rad).

ChIP-exo and deep sequencing

E. coli and *Salmonella* strains carrying biologically functional UvrY-FLAG [4] or SirA-FLAG proteins (S1 Fig) were grown in Kornberg medium and LB (supplemented with 10 mM glucose), respectively at 37°C, 250 rpm. At mid-exponential phase of growth (OD$_{600}$ of 0.6), formaldehyde (1% final concentration) was added and the cultures were incubated for 20 min at 30°C, 150 rpm. The crosslinking reaction was then quenched by the addition of 5 mL of 1.0 M glycine, pH 8.0 to 10 mL of culture. The samples were kept at room temperature for 5 min with gentle swirling. The cells were harvested by centrifugation (6000 x G, 5 min, 4°C), washed twice with ice-cold 1x PBS, suspended in 500 μL lysis buffer (10 mM Tris-HCl pH 8.8, 50 mM NaCl, 20% [w/v] sucrose, 10 mM EDTA) containing protease inhibitor cocktail (cOmplete Mini, EDTA-free protease inhibitor cocktail, Roche Diagnostics) and 2 mg/mL lysozyme. After 30 min on ice, 500 μl of 2x IP buffer (100 mM Tris-HCl, pH 7.0, 300 mM NaCl, 2% Triton X-100, 2 mM EDTA) was added and the samples were incubated at 37°C for 10 min, followed by 2 min on ice. The samples were then sonicated on ice, (60 pulses total, 5 sec on 5 sec off per pulse, in a total of 6 sets, separated by 2 min on ice, using a sonicator (Fisher Scientific, Sonic Dismembrator Model 500, set at 20% amplitude). Unbroken cells and debris were removed by centrifugation. Sonicated samples (1 mL) were treated with 20 μL ANTI-FLAG® M2 beads (Sigma) by spinning on a tube rotating mixer (Thermo Scientific) at 4°C for 4 hr to immunoprecipitate UvrY-FLAG. The beads were washed three times with 1mL 1X IP buffer and two times with 1mL 1X TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The beads were resuspended in 100 μL TE buffer and the formaldehyde crosslinking was reversed by heating at 95°C for 20 min. Afterward, the samples were incubated with 8 μL of 10 mg/mL RNase A for 2 h at 37°C and with 4 μL of 20 mg/mL proteinase K at 55°C for 2h. The resulting DNA was purified using a Qiagen MinElute PCR Purification Kit according to manufacturer's instructions. As a control, wild type strains expressing no FLAG epitope tagged protein were treated using the same procedures. The specificity of the ChIP assay was confirmed by amplifying promoter DNA of the *csrB*, *lacY* and 16s rRNA (*rrsH*) genes with primers (S1 Table) using the DNA recovered from the ChIP reactions as a template (S2 Fig). *csrB* was used as a positive control for UvrY/SirA binding while *lacY* (*E. coli*) and 16S rDNA (*Salmonella*) were used as negative controls. After confirming the specificity of the ChIP reactions, the DNA samples were frozen and shipped to Peconic LLC for the remainder of the ChIP-exo procedure and deep sequencing with Illumina Hi-seq 2000 as described previously [34].

ChIP-exo data analysis

Sequencing reads were mapped to their respective *E. coli* MG1655 (NC_000913.3) and *Salmonella enterica* subsp. *enterica* serovar Typhimurium 14028S (NC_016856.1) genomes with BWA (Burrows-Wheeler Aligner, available from http://bio-bwa.sourceforge.net). Alignments were visualized using the Integrated Genome Viewer (IGV, Broad Institute) [44, 45]. Peaks were identified in the mapped reads with GeneTrack [46]. Low confidence singleton peaks and those without matching peaks on the opposing strand were discarded. Enriched genomic loci were classified into three groups based on enrichment of read counts relative to read counts within 500 bp upstream of the transcription start sites of *lacY* and/or 16s rRNA (*rrsH*), which were determined not to be direct targets of UvrY/SirA (S2 Fig). Maps of enriched regions are shown as bedgraphs, which display in a continuous fashion the abundance of reads identified at particular position in the genome from the sequencing analysis. For motif discovery, DNA

sequences from the peak-pair midpoint of each of the enriched genomic loci, with +/- 40nt upstream and downstream extension distance were extracted. Centered in the 81nt-long extracted DNA sequence is the UvrY/SirA crosslinking site. The extracted DNA sequences were uploaded into MEME Suite [47] for motif discovery analysis. The datasets from ChIP-exo analyses were submitted to the NCBI GEO repository under the accession number GSE74810.

RNA extraction and Northern blotting

Procedures for RNA extraction and Northern blotting were conducted as described previously, with minor modification [4]. Briefly, bacteria were harvested at mid-exponential phase (OD_{600} of 0.6) or as otherwise stated and total cellular RNA was prepared (RNeasy mini kit, Qiagen). RNA was separated by electrophoresis on 5% polyacrylamide gels containing 7 M urea, transferred to a positively charged nylon membrane (Roche Diagnostics) by electro-blotting, and crosslinked to the membrane with UV light. Crosslinked RNA was hybridized to DIG-labeled antisense probe (68°C, overnight) and signal was developed using the DIG Northern Starter kit (Roche Diagnostics). Antisense RNAs were prepared from PCR products using the DIG Northern Starter kit (Roche Diagnostics). Blots were imaged using a ChemiDoc XRS+ system (Bio-Rad) and RNA signals quantified using Quantity One image analysis software (Bio-Rad). The 5S or 16S and 23S rRNAs served as loading controls, and were detected by hybridization or methylene blue staining, respectively.

Western blotting

Western blotting was conducted as previously described [4]. Briefly, proteins (10 μg) were separated using SDS-PAGE and electroblotted onto 0.2 mm polyvinylidene difluoride membranes.Anti-FLAG® M2 monoclonal antibody (Sigma) and anti-RpoB monoclonal antibody (Neoclone) were used for detection of the FLAG epitope and RpoB. Signal detection used treatment with horseradish peroxidase-linked secondary antibodies followed by SuperSignal® West Femto Chemiluminescent Substrate (Thermo Scientific). Blots were imaged using the ChemiDoc XRS+ system (Bio-Rad) and the signals were quantified using Quantity One image analysis software (Bio-Rad).

Analysis of UvrY-FLAG protein stability

Bacterial cells were grown to mid-exponential phase of growth, at which point protein synthesis was stopped by the addition of tetracycline (200 μg/mL) and chloramphenicol (100 μg/mL). Cells were collected at several times following the addition of the antibiotics, harvested by centrifugation and were immediately mixed with Laemmli sample buffer and lysed by sonication and boiling. Samples (10 μg protein) were subjected to SDS- PAGE and were analyzed by Western blotting as described above.

β-Galactosidase assays

The assays for specificβ-Galactosidase activity were conducted as described previously [2]. Values represent the averages from two independent experiments. Error bars represent standard errors of the means.

Analysis of UvrY phosphorylation in vivo

Studies to determine the extent of phosphorylation of the UvrY-FLAG protein in vivo were conducted as previously described [4]. Briefly, proteins from cell lysates were fractionated on

Phos-tag^{TM} SDS-PAGE gels, which resolve UvrY from UvrY-P, and the FLAG epitope of the UvrY-FLAG protein was detected by Western blotting.

Quantitative RT-PCR

Quantitative Real Time-PCR (q-RT-PCR) was conducted in an iCycler™ thermocycler (Bio-Rad) using the iScript™ One-Step RT-PCR Kit with SYBR® Green (Bio- Rad) as described previously, with minor modifications [4]. Reactions (15 µl) contained immunoprecipitated DNA (50 ng/µl), primers (0.5 µM each; S1 Table) and SYBR® Green RT-PCR Reaction Mix. PCR cycle parameters were as follows: 40 cycles of PCR at 95°C denaturation for 10 s, 60°C of annealing, extension, and detection for 30 s. The specificity of the PCR product was determined by melting curve analysis with reference to the calculated Tm. For melting curve analysis, the temperature was increased from 60°C to 95°C at a rate of 0.5°C/10 s. PCR product concentration was determined using a standard curve prepared with iCycler iQ optical system software version 3.1 (Bio-Rad), according to the manufacturer's instructions.

Analysis of UvrY binding to *csrB* DNA *in vivo* by ChIP-PCR

UvrY binding to *csrB* genomic DNA *in vivo* was analyzed by ChIP-quantitative PCR (ChIP-PCR). The first steps of the ChIP-PCR assay were conducted as described for the ChIP-exo assay. After releasing the formaldehyde crosslinks, the DNA was extracted and purified using Qiagen MinElute Purification Kit. The isolated *csrB* DNA was assayed using quantitative real time-PCR (q-RT-PCR) with primers listed in (S1 Table). The *lacY* gene served as a negative control for this reaction using primers also listed in (S1 Table).

Electrophoretic gel mobility shift assay (EMSA) for DNA binding

For DNA electrophoretic gel mobility shift assays, the regions from -246 to +49 and -247 to +56, with respect to the transcriptional start sites of *csrB* and *csrC*, respectively, were amplified by PCR and end-labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase. Binding reactions (10 µl) contained 0.5 nM of end labeled DNA, 20 mM Tris HCl (pH 7.5), 10% (v/v) Glycerol, 50 mM KCl, 3 mM MgCl$_2$, 1 mM dithiothreitol, 100 µg/mL BSA and phosphorylated or non-phosphorylated UvrY-His$_6$ protein. Reactions were incubated for 30 min at 37°C, then 1µl xylene cyanol was added and samples were separated by electrophoresis on 7% non-denaturing polyacrylamide gels with 0.5X TBE as the running buffer. After electrophoresis, the gels were vacuum dried, the radioactive signals captured by phosphorimage analysis (PMI™, Bio-Rad) and analyzed using Quantity One software.

DNase I footprinting

Double stranded DNA for footprinting of *csrB* (466 bp long, extending from 420 bp upstream to 46 bp downstream of the TSS) or *csrC* (419 bp long, extending from 319 bp upstream to 100 bp downstream of the TSS) was prepared by PCR using strand-specific [^{32}P] 5'-end-primers (S1 Table). The 5'-end-labeled PCR products were gel-purified and the DNA (0.5 nM) was used for binding reactions (10 µL) containing 20 mM Tris HCl (pH 7.5), 10% (v/v) glycerol, 50 mM KCl, 3 mM MgCl$_2$, 1mM dithiothreitol and 100 µg/mL bovine serum albumen along with phosphorylated or non-phosphorylated UvrY-His$_6$ protein. The binding reactions were incubated at 37°C degrees for 30 min followed by treatment with 0.025 U DNase I (Roche) per reaction, for 1 min at 37°C. The DNase I was inactivated by heating at 75°C degrees for 10 min. Reactions were then mixed with formamide loading buffer (10 mL formamide, 10 mg xylene cyanol FF and 10 mg bromophenol blue), denatured at 95°C for 5 min, and separated by

electrophoresis on a 6% denaturing polyacrylamide gel. The gel was vacuum dried, radioactive signals captured by phosphor imaging (PMI™, Bio-Rad) and analyzed with Quantity One software.

Construction of plasmid-borne *csrB-lacZ* and *csrC-lacZ* fusions for S-30 transcription-translation

The *csrB-lacZ* and *csrC-lacZ* carrying plasmids, pLFXcsrB-lacZ and pLFXcsrC-lacZ, were constructed by first amplifying the 502 nt (-500 to +2 with respect to *csrB* transcriptional start site) and 304 nt (-301 to +3 with respect to *csrC* transcriptional start site) genomic regions using the primer pairs csrB lacZ Fwd / csrB lacZ Rev and csrC lacZ Fwd / csrC lacZ Rev (S1 Table). The PCR products were gel purified, digested with *Pst*I and *Kpn*I, ligated to *Pst*I- and *Kpn*I-digested and dephosphorylated plasmid pLFX, and electroporated into DH5αλpir cells. DNA sequencing was used to confirm that the cloned regions did not contain any mutations.

In vitro coupled transcription-translation

In vitro transcription-translational assays used pLFXcsrB-lacZ and pLFXcsrC-lacZ supercoiled plasmids and were performed with S-30 extracts prepared from the UvrY deficient strain, CF7789 *uvrY::cam*, as described previously [16], except that reactions (32 μl) contained 0.5 U *E. coli* RNA polymerase holoenzyme and 3 μl of ^{35}S-methionine (1175 Ci/mmol). The UvrY-P (2.3 μM), ppGpp (250 μM) and DksA (2 μM) were included, as indicated, in the reactions. IHF (1μM) was added to all of the reactions that contained pLFXcsrB-lacZ plasmid. DksA protein was a generous gift of Prof. Richard Gourse, University of Wisconsin, Madison. IHF protein was a generous gift from Prof. Anca Segall, San Diego State University. Incorporation of ^{35}S-methionine into protein products was determined after SDS PAGE separation by using phosphorimaging. Signal intensity was determined using Quantity One software.

Phylogenetic analysis of BarA, UvrY, CsrA, and FliW distribution

We identified CsrA, BarA, UvrY and FliW orthologs from all fully-sequenced bacterial genomes in the NCBI genomes database (http://www.ncbi.nlm.nih.gov/genome/), using the NCBI Prokaryotic Genome Annotation Pipeline v2.0 to assign orthology [48]. This pipeline combines a sequence similarity-based approach with the comparison of the predicted gene products to the nonredundant protein database, Entrez Protein Clusters, the Conserved Domain Database (CDD) [48].

In order to filter out false positives and false negatives from the NCBI orthologs data set, we aligned a set of representative sequences against the genomes with predicted orthologs using tblastn [49]. The alignment results were filtered by sequence similarity and alignment coverage: BarA was filtered by 50% similarity and 50% coverage, UvrY was filtered by 75% similarity and 80% coverage, CsrA was filtered by 60% similarity and 50% coverage, and FliW was filtered by 50% similarity and 50% coverage. All the similarity and coverage thresholds were established by the minimum similarity and coverage observed between annotated sequences from NCBI nucleotide database and the following set of representatives used as query sequences: BarA from *Escherichia coli* (NCBI protein identification number PI: 190908466), UvrY from *E. coli* (PI: 190906390), CsrA from *E. coli* and *Bacillus subtilis* (PI: 257755450 and 459391362), and FliW from *Bacillus amyloliquefaciens* (PI: 307608134). After filtering the results by alignment similarity and coverage, we performed a manual refinement of the results by verifying the pairwise alignments, sequence annotations, and bibliographic information.

A reference phylogeny of fully-sequenced genomes encoding CsrA and at least one of BarA, UvrY or FliW was extracted from the NCBI taxonomy database [50] using NCBI-taxcollector

[51] and PhyloT v2015.1 (http://phylot.biobyte.de), and gene presence/absence data were visualized using iTOL v3.0 [52]. Correlations among presence/absence of CsrA, BarA, UvrY and FliW were calculated using Pearson's product moment correlation, and significance was assessed using Fisher's Z transform (implemented in the R stats package, v3.2.0). For calculating the correlations, we clustered the results at species and genus level in order to minimize the number of false negatives and false positives. For clustering the results at species level, we established the presence/absence as the value that appears most often at strain level (statistical mode). For clustering the results at genus level, we established the presence/absence as the mean calculated at species level.

In order to determine if any strong biases in the methodology used to identify orthologous genes could affect our results we used two additional and widely-employed methods to identify alternative ortholog sets: KEGG orthology database (KO) [53], and UniProt reference clusters of orthologs (UniRef) [54]. KEGG database defines the orthologs by comparing experimental data with KEGG pathway maps, BRITE functional hierarchies and KEGG modules. UniProt defines the orthologs by clustering protein sequences by 50% identity. We compared the correlations calculated for these alternative approaches to the results obtained from the NCBI Prokaryotic Genome Annotation Pipeline, at genus and species level.

Results and Discussion

Determination of genome-wide UvrY/SirA-binding loci using ChIP-exo

To probe for UvrY/SirA DNA binding sites in *E. coli* and *Salmonella*, we used ChIP-exo, a comprehensive genomic DNA-protein interaction assay with single nucleotide resolution and high specificity [34]. This assay uses nuclease trimming reactions to remove nonspecific DNA from the immunoprecipitated complexes and to identify the crosslinking sites much more precisely than ChIP-seq [34]. The experiments were performed with mid-exponential phase cultures (OD$_{600}$ of 0.6) of strains carrying FLAG-tagged UvrY or SirA proteins, respectively, expressed from the native genomic loci. Altogether, 44 million sequencing reads for UvrY (*E. coli*) and 31 million for SirA (*Salmonella*) were mapped to their respective genomes. From these analyses, two highly enriched genomic loci were identified at the *csrB* and *csrC* genes of both *E. coli* (Fig 1A and 1B) and *Salmonella* (Fig 2A and 2B). These results indicated that the *csrB* and *csrC* genes represent the strongest targets of UvrY/SirA binding in these species.

In addition to the *csrB* and *csrC* genes, weakly enriched genomic loci were identified proximal to the promoter regions of 286 genes in *E. coli* and 301 genes in *Salmonella* (S2 Table). These enriched genomic loci were arbitrarily classified into three groups based on enrichment of their occupancy peaks over background (S2 Table). Accordingly, genomic loci enriched 5-fold or greater than the *lacY* and/or 16s rRNA (*rrsH*) genes were included in group one. Besides *csrB* and *csrC*, the promoter regions of *fhuF* and *spf* from *E. coli* and *spf* from *Salmonella* fell into group one. Genomic loci that were enriched 2- to 5-fold comprised group two, which included the promoter regions of 9 genes in *E. coli* and 8 genes in *Salmonella*. The rest of the genomic loci, which include regions proximal to the promoter regions of 275 genes in *E. coli* and 292 genes in *Salmonella*, have only 1.5- to 2.0 occupancy read fold over that of *lacY* or 16s rDNA (*rrsH*) and were grouped into group three.

UvrY/SirA consensus DNA binding motif

Further inspection of the ChIP-exo results showed enriched crosslinking in the upstream locations of *csrB* and *csrC*, with sequences extending from -222 to -142 from the start of transcription (*csrB*) or -223 to -143 (*csrC*) in *E. coli* (Fig 1D and 1E) and from -215 to -135 (*csrB*) and -181 to -101 (*csrC*) in *Salmonella* (Fig 2C and 2D). Distinct sites of crosslinking in these genes
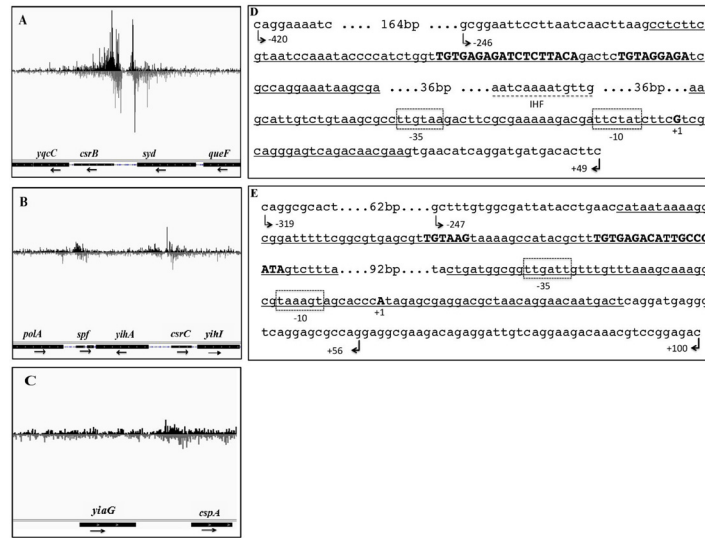
Fig 1. Genomic binding sites for UvrY identified by ChIP-exo. Bedgraphs depict genomic loci enriched by co-immunoprecipitation with cross-linked UvrY-FLAG, proximal to the *csrB* (A), *csrC* (B), *spf* (B) and *cspA* (C) genes. Two UvrY crosslinking sites were discovered in the promoter regions of *csrB* (panel A, underlined in panel D) and *csrC* (panel B, underlined in panel E). One site is within the region extending from -222 to -142 in *csrB* or from -223 to -143 in *csrC*. The other crosslinking site lies close to the promoter, extending from -56 to +25 in *csrB* and from -49 to +32 in *csrC* promoter regions. A 9 bp inverted repeat DNA sequence within the upstream crosslinking regions is bolded and capitalized. A putative IHF binding site (broken underline) is located between the two UvrY crosslinking sites in the promoter region of *csrB* (D), but was not apparent in *csrC*. The DNA used for DNase I footprinting included 466 bp (-420 to +46) for *csrB* and 419 bp (-319 to +100) for *csrC*. DNA fragments used for electrophoretic mobility shift assay included (-246 to +49) for *csrB* and (-247 to +56) for *csrC*. The images were constructed using the Integrated Genome Viewer (IGV, Broad Institute) (44, 45).

doi:10.1371/journal.pone.0145035.g001

were also observed at downstream sequences that extended from -56 to +25 (*csrB*) and from -49 to +32 (*csrC*) in *E. coli* (Fig 1D and 1E) and from -41 to +40 (*csrB*) and from -81 to -1 (*csrC*) in *Salmonella* (Fig 2C and 2D). In the center of each of the upstream putative binding
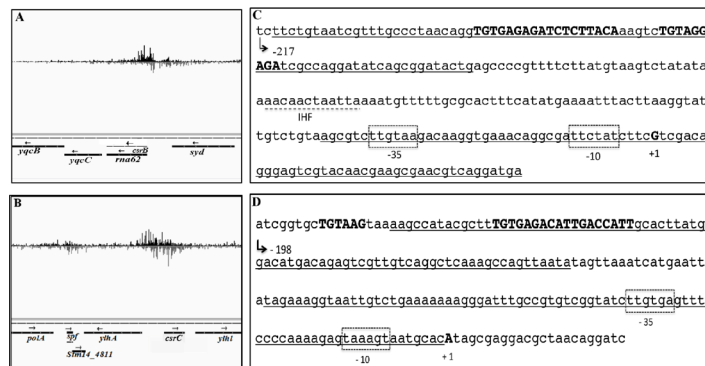

Fig 2. Genomic binding sites for SirA discovered by ChIP-exo. Bedgraphs show enriched genomic loci, proximal to the regulatory regions of *csrB* (A), *csrC* (B) and *spf* (B). Two putative SirA binding sites are shown for the promoter regions of *csrB* (*rna62*) (panel A, underlined in panel C) and *csrC* (panel B, underlined in panel D). One site is located in the region from -215 to -135 in *csrB* and from -181 to -101 *csrC* promoter regions, respectively. The other putative binding site is close to the promoter, within the -41 to +40 (*csrB*) and -81 to -1 in *csrC* promoter regions, respectively. A 9 bp-long inverted repeat DNA sequence in the upstream sites is shown in bold and capitalized. A putative IHF binding site is marked between the two putative UvrY binding sites in the promoter region of *csrB* (C, broken underline), but was not apparent in *csrC*. The images were constructed using the Integrated Genome Viewer (IGV, Broad Institute) (44, 45).

doi:10.1371/journal.pone.0145035.g002

sites is an 18 nt nearly perfect palindrome or inverted repeat sequence (IR). In *csrB*, the sequence, TGTGAGAGATCTCTTACA, is centered at -183/-182 in *E. coli* ([Fig 1D](#)) and -182/-181 in *Salmonella* ([Fig 2C](#)). Moreover, a partially conserved additional sequence (TGTAG-GAGA) located 5 bp downstream of the IR, is seen in the *csrB* promoter of *E. coli* ([Fig 1D](#)) and *Salmonella* ([Fig 2C](#)). The IR of *csrC* shows weaker symmetry in both *E. coli* (TGTGAGA-CATTGCCGATA) ([Fig 1E](#)) and *Salmonella* (TGTGAGACATTGACCATT) ([Fig 2D](#)). Moreover, a partially conserved sequence (TGTAAG) representing half of the IR, located 16 bp upstream of the IR, is also seen in the *csrC* promoter of *E. coli* ([Fig 1E](#)) and *Salmonella* ([Fig 2D](#)). In contrast, the IR is not preserved in the downstream crosslinking sites of the *csrB* or *csrC* genes.

To search for the conserved IR at the weaker targets of UvrY/SirA ([S2 Table](#)), DNA sequences from these enriched promoter regions were analyzed using MEME Suite software [47]. These analyses showed that the IR sequence is not conserved in any of the weaker putative binding sites.

UvrY requires phosphorylation for DNA binding

In the BarA-UvrY TCS, BarA is the histidine kinase, which upon signal detection, autophosphorylates and transfers the phosphoryl group to a conserved aspartate residue of UvrY [20]. The phosphorylated UvrY is then thought to bind specific DNA sequences in the promoters of its target genes and regulate their transcription [18, 20, 22]. MBP-SirA was previously found to bind similarly to *csrB* DNA *in vitro* regardless of whether it had been phosphorylated or not, starting at a concentration of 1.5 μM [25], suggesting that MBP-SirA phosphorylation might not be required for DNA binding *in vitro*. This observation was consistent with another report, showing that phosphorylation only increased the DNA binding affinity of SirA-His$_6$ by approximately two-fold *in vitro* [19].

In contrast to results from *in vitro* DNA binding experiments, deletion of the gene for BarA sensor kinase caused more than a 10-fold decrease in the level of CsrB RNA [16, 22, 35]. Similarly, in *Salmonella*, substitution of alanine for the predicted phosphorylated histine residue of SirA, Asp54, caused the loss of CsrB expression [25]. Thus, it appears that phosphorylation is critical for UvrY activity *in vivo*. Whether UvrY requires phosphorylation for efficient DNA binding *in vivo* or for later steps in transcription initiation was not clear.

To address this issue, we used both *in vitro* and *in vivo* assays to investigate the role of UvrY phosphorylation in DNA binding. We performed electrophoretic mobility shift assay (EMSA) using phosphorylated (UvrY-P) and non-phosphorylated forms of the recombinant protein, UvrY-His$_6$, which was determined to be functional *in vivo* ([S1 Fig](#)). DNA fragments that encompass the ChIP-exo-derived putative UvrY binding sites in the promoter regions of *csrB* ([Fig 1D](#)) and *csrC* ([Fig 1E](#)) were used as probes. Unlabeled *csrB* and *rrlE* DNA fragments were used as competitive and non-competitive DNA probes, respectively. Similar to the previous observations [19, 25], our results showed that *in vitro* phosphorylation led to a modest (~2-fold) increase in DNA binding affinity of UvrY to *csrB* ([Fig 3A and 3B](#)) and *csrC* DNA ([Fig 3C](#)), as compared to the non-phosphorylated UvrY. However, we also observed that the UvrY protein preparation contained around 7% of UvrY-P prior to the *in vitro* phosphorylation reaction ([Fig 3D](#)). It is possible that this contaminating UvrY-P was entirely responsible for binding to *csrB* and *csrC* DNA *in vitro* ([Fig 3B](#)).

To determine whether UvrY requires phosphorylation for *in vivo* DNA binding, we tested UvrY-FLAG binding to *csrB* *in vivo* in a strain that lacked BarA (ΔbarA) and its isogenic *barA* wild-type strain. For this analysis, we first determined the phosphorylation state of UvrY-FLAG protein in the presence and absence of BarA. The results indicated that in the presence
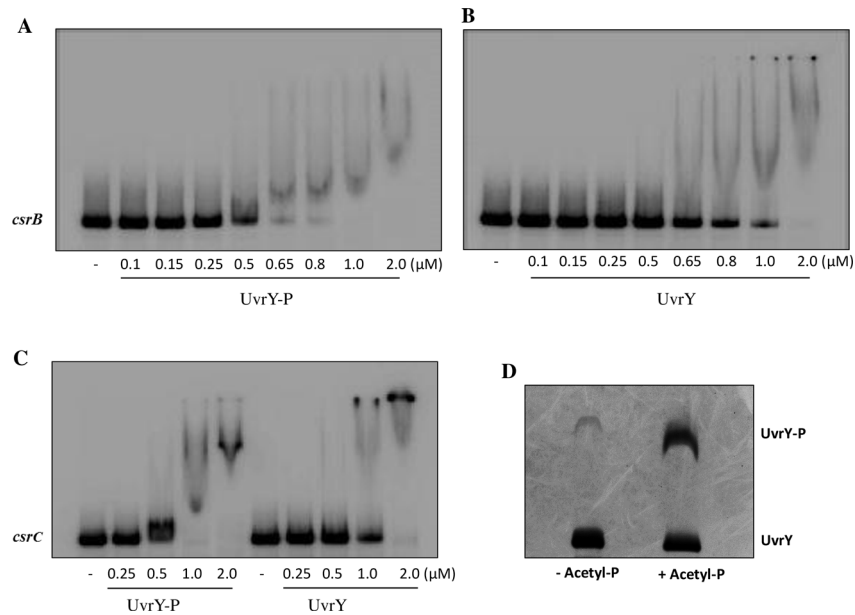
Fig 3. Electrophoretic gel mobility shift assay showing UvrY binding to *csrB* and *csrC* DNA. Binding of phosphorylated (UvrY-P) and non-phosphorylated (UvrY) UvrY-His$_6$ to *csrB* DNA (A and B) and *csrC* DNA (C) was tested as shown. The *csrB* and *csrC* DNA probes (0.5 nM) used for this experiment (depicted in Fig 1D and 1E, respectively) were incubated with increasing concentrations of *in vitro* phosphorylated or non-phosphorylated UvrY-His6 protein for 30 min at room temperature. The DNA-protein complexes were resolved by electrophoresis on a non-denaturing 7% polyacrylamide gel. The phosphorylation state of the UvrY-His$_6$ protein used in these experiments was determined by Phos-tag SDS PAGE gel analysis (D).

doi:10.1371/journal.pone.0145035.g003

of BarA, approximately 7% of the UvrY protein was phosphorylated (S3 Fig), similar to a previous determination [4]. In the absence of BarA, no detectable phosphorylation of UvrY-FLAG protein was observed in LB medium at mid-exponential phase of growth (S3 Fig). We next used ChIP-PCR to test for *in vivo* binding of UvrY to *csrB* DNA under this growth condition. In this experiment, an approximate 35-fold reduction in DNA binding was observed in the strain that lacked BarA (Δ*barA*) relative to the isogenic *barA* wild-type strain (Fig 4). These results show for the first time that UvrY requires phosphorylation for effective DNA binding *in vivo*.
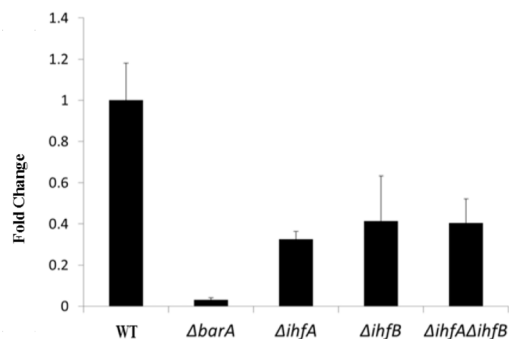

Fig 4. Effects of BarA and IHF on *in vivo* binding of UvrY to *csrB*. The effects of BarA, IhfA and IhfB on *in vivo* binding of UvrY to *csrB* promoter were determined by ChIP-quantitative PCR (ChIP-PCR) in a WT strain (MG1655 expressing UvrY-FLAG) and isogenic Δ*barA*, Δ*ihfA*, Δ*ihfB* and Δ*ihfA*Δ*ihfB* strains, as described in the Experimental Procedures. The *lacY* gene served as a negative control for this reaction. Data depict the results of three independent experiments. Error bars represent the standard errors of the means.
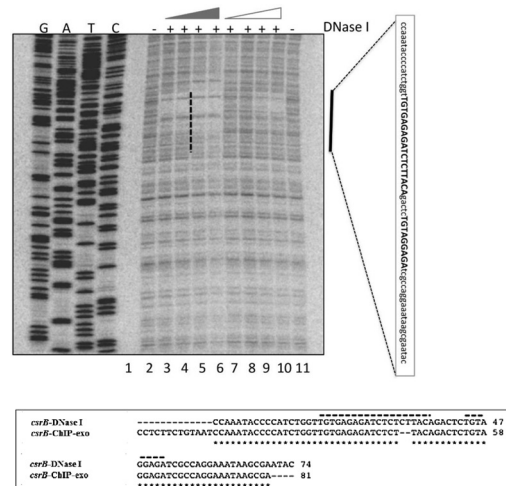
doi:10.1371/journal.pone.0145035.g004

Fig 5. DNase I footprinting of *E. coli csrB* DNA using phosphorylated and non-phosphorylated UvrY. A ^{32}P-end labeled DNA probe that included both the upstream and downstream putative UvrY binding sites was used for these experiments (shown in Fig 1D). Reactions in all lanes except 1 contained DNase I (0.025U/ 12.5ul reaction). Reactions in lanes 3–6 and lanes 7–10 contained 0.25, 0.35, 0.5, 0.7 µM of phosphorylated or non-phosphorylated UvrY-His$_6$, respectively. Lane 2 reaction contained no UvrY protein. A vertical black bar indicates a protected region, and the sequence corresponding to the protected region is shown in a vertical rectangular box. An alignment of sequences corresponding to the protected regions from DNase I and the ChIP-exo results is shown in the horizontal rectangular box. The 18nt-long palindromic sequence and the partially conserved palindromic sequences are marked with broken black lines.

doi:10.1371/journal.pone.0145035.g005

UvrY specifically binds to an 18 nt-long IR DNA sequence

In order to more precisely define the UvrY DNA binding sites, we performed *in vitro* DNase I footprinting experiments on *csrB* and *csrC* DNA using phosphorylated and non-phosphory-lated UvrY-His$_6$ protein. The DNA probes for this experiment encompassed both the upstream and downstream putative binding sites of these genes (Fig 1D and 1E). The results of these experiments showed protection of only the upstream binding sites containing the IR sequences of both *csrB* (Fig 5) and *csrC* (Fig 6 and S4 Fig). The putative downstream binding sites that were observed *in vivo* (Fig 1A and 1B & 1D and 1E) were not protected *in vitro*. In addition, only the phosphorylated UvrY-His6 was observed to protect the IR, which further indicates that UvrY requires phosphorylation for tight, specific binding. Together, these results suggest that UvrY-P binds specifically and directly to the IR sequence of these genes. However, *in vivo* binding of UvrY-P to the downstream sequences of these genes either requires conditions or factor(s) that were absent from the footprinting reactions or perhaps more likely, UvrY-P becomes cross-linked to DNA indirectly through interactions with DNA binding proteins such as RNA polymerase, which must bind to the downstream regions of these genes in order to initiate transcription.

UvrY requires IHF for optimal binding to and expression of *csrB* but not *csrC*

The nucleoid-associated protein Integration Host Factor (IHF) is a heterodimeric protein composed of two homologous subunits, IHFα (IhfA) and IHFβ (IhfB), which facilitates the transcription of many genes by bending the DNA [55–57]. In the Csr/Rsm system, IHF was shown to directly bind to the promoter of the *csrB* gene of *Salmonella*) [18] and the *rsmZ* promoter of *Pseudomonas fluorescens* [58]. In *Salmonella*, deletion of *ihfA* was also shown to decrease *csrB*
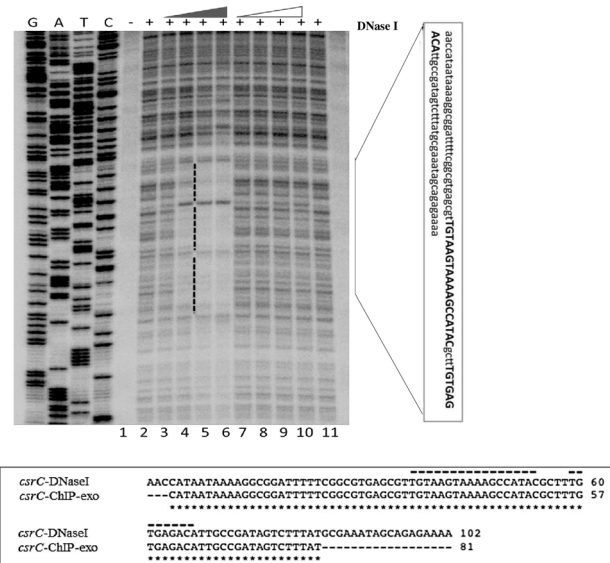
Fig 6. DNase I footprinting of *E. coli csrC* DNA by phosphorylated and non-phosphorylated UvrY. A
[32]P-end labeled DNA probe that included both the upstream and downstream putative UvrY binding sites was
used for these experiments (Fig 1E). Reactions in all lanes except lanes 1 contained DNase I (0.025U/12.5ul
reaction). Reactions in lanes 3–6 and lanes 7–10 contained 0.25, 0.35, 0.5, 0.7 μM of phosphorylated or non-
phosphorylated UvrY-His$_6$, respectively. Lane 2 reaction contained no UvrY. A vertical black bar indicates a
protected region, and the sequence corresponding to the protected region is shown in a vertical rectangular
box. An alignment of sequences corresponding to the protected regions from DNase I and ChIP-exo results is
shown in the horizontal rectangular box. The 18nt-long palindromic sequence and the partially conserved
palindromic sequences are marked with broken black lines.

doi:10.1371/journal.pone.0145035.g006

expression [18]. However, whether the observed effect of IHF in sRNA expression in those bac-
terial species is UvrY-mediated and if so, whether IHF is required for UvrY to bind to DNA or
for later transcription initiation steps is not clear.

In this study, we identified a putative IHF binding site in the promoter region of *csrB* of *E.
coli* (Fig 1D) and *Salmonella* (Fig 2C), in agreement with an earlier report [18], located between
the upstream and downstream *in vivo* crosslinking sites for UvrY-P, but we did not identify a
similar site within the *csrC* genes. To determine whether IHF affects expression of *csrB* in *E.
coli*, we measured the levels of CsrB in the presence and absence of IHF (Δ*ihfA* and/or Δ*ihfB*).
Our results revealed a ~10-fold reduction in CsrB RNA levels in the absence of IhfA, IhfB, or
both, compared to the isogenic wild type strain (Fig 7A). We also performed epistasis experi-
ments in which we measured the effect of IHF (Δ*ihfA* and/or Δ*ihfB*) on the expression of *csrB*
in a strain lacking *uvrY* (Δ*uvrY*). Our results showed that neither IhfA nor IhfB affected the lev-
els of CsrB in the *uvrY* mutant strain (Fig 7A), suggesting that UvrY mediates the effect of IHF
on the expression of *csrB*. Analysis of *csrB* DNA binding by UvrY-FLAG using *in vivo*
ChIP-PCR analyses showed that binding was reduced by approximately 2.5-fold in the absence
of either IhfA or IhfB (Fig 4), suggesting that IHF is required for optimal binding of UvrY-P to
csrB DNA *in vivo*. However, this modest effect of IHF on UvrY DNA binding does not appear
to account for its 10 to 12-fold effect on *csrB* expression. Therefore, IHF also appears to affect
later steps in *csrB* transcription (Fig 7A). Western blotting experiments demonstrated that IHF
did not affect the phosphorylation (S3 Fig) or expression of UvrY-FLAG (S5 Fig). Hence, our
results reveal that IHF is required for optimum binding of UvrY-P to *csrB* DNA and activation
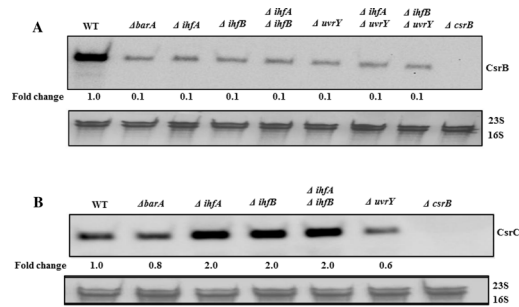of *csrB* transcription.

Fig 7. Effects of UvrY, BarA and IHF on CsrB/C sRNAs levels. Northern blots showing effect of several gene deletions on the levels of *E. coli* CsrB (A) and CsrC (B). Cultures were grown in LB to mid-exponential growth phase (OD$_{600}$ of 0.6). The 16S/23S rRNA loading controls are also shown.

doi:10.1371/journal.pone.0145035.g007

IHF facilitates the transcription of many genes by bringing relatively distant DNA sites (~200 bp) closer together in space by bending the DNA (~140°) [59, 60]. Therefore, we propose that UvrY-P bound at the upstream IR site (centered at -183/-182 in *csrB* of *E. coli* and at -182/-181 in the *csrB* of *Salmonella*) is brought into the proximity of the *csrB* promoter-RNA polymerase complex by IHF-mediated bending of the DNA. Such DNA bending may lead to UvrY-P binding to the downstream DNA crosslinking site directly or indirectly through interactions with RNA polymerase or perhaps with other unknown DNA binding factor(s). This type of transcriptional activation, known as repositioning, is observed in promoters where the primary activator is unable to make a productive contact with RNA polymerase without being repositioned by a secondary activator [60, 61]. Such a mechanism appears to be required for the activation of the *E. coli narG* promoter by NarL, another FixJ family response regulator, which binds at –190, while IHF binds at -125 and Fnr binds at – 41 of *narG* [59, 60].

In contrast to its effect on *csrB*, IHF was not needed for expression of *csrC* (Fig 7B). In fact, in the single and double IHF deletion strains, CsrC RNA levels were 2-fold higher than in the isogenic wild type strain (Fig 7B). This may be because in the Δ*ihfA* Δ*ihfB* strains, the lower levels of CsrB RNA (Fig 7A) lead to an increase in the concentration of free CsrA in the cell, which in turn leads to increased levels of CsrC via a negative feedback loop in the Csr system [13].

UvrY effects on the expression of other putative *in vivo* targets

To examine the regulatory effects of UvrY on the expression of the potential new target genes discovered by ChIP-exo (S2 Table), we tested several gene products and transcripts from groups one, two and three by Western and Northern blotting. Our results showed little or no regulatory effects of UvrY on the expression of the genes that were tested (S6A and S6C Fig), with the exception of *cspA*. UvrY showed a relatively modest, but reproducible negative effect on the expression of the cold-shock protein CspA (Fig 8).

A 2-fold increase in the levels of CspA-FLAG was observed in a Δ*uvrY* strain compared to the isogenic wild type strain (Fig 8). Furthermore, disruption of *csrA* caused a modest decrease in CspA-FLAG levels, suggesting that CsrA may activate *cspA* expression. To determine whether UvrY regulates *cspA* indirectly through its effects on CsrB/C, and therefore CsrA activity, we conducted an epistasis analysis (Fig 8). A strain that was disrupted in both *uvrY* and *csrA* was transformed with plasmids expressing *uvrY* or *csrA* from the plasmid cloning vector pBR322, and CspA-FLAG protein levels were determined in the resulting strains. We observed that CspA-FLAG levels were essentially identical in strains containing pBR322 and the *csrA*-expression plasmid pCRA16. In contrast, the *uvrY*-expression plasmid (pUY14) led to modest
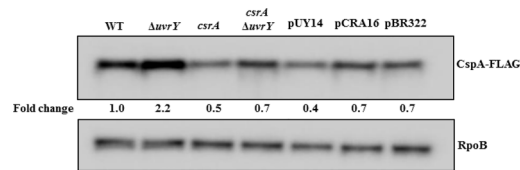
Fig 8. Effects of genes encoding UvrY and Csr factors on CspA protein levels. Western blot showing effects of *uvrY* deletion and *csrA*::kan disruption on the levels of CspA-FLAG protein in an MG1655 derivative expressing CspA-FLAG (WT) from the *cspA* genomic locus. Effect of *csrA*::kan Δ*uvrY* complemented with UvrY (pUY14), CsrA (pCRA16) or control (pBR322) expression plasmids is also shown. Cultures were grown at 37°C in LB to mid-exponential growth phase (OD$_{600}$ of 0.6). The RpoB protein served as a loading control. This experiment was repeated at least three times with reproducible results.

doi:10.1371/journal.pone.0145035.g008

decrease in CspA-FLAG levels in this strain. These results are consistent with a model in which UvrY inhibits *cspA* expression by binding to the *cspA* promoter, while the effect of CsrA may be mediated indirectly. Also consistent with this model, *in silico* analyses did not reveal typical CsrA binding sequences (GGA) in the 160 NT 5'-UTR of the *cspA* transcript (data not shown) and an RNA-seq analysis conducted previously did not reveal *cspA* mRNA among the 721 different transcripts that copurified with CsrA [2].

Based on all of these observations, we conclude that, under the growth conditions tested, the BarA-UvrY TCS directly exerts its global effect on gene expression primarily through the Csr system by activating the transcription of the CsrB and CsrC sRNAs, with *cspA* representing a likely exception. This raises the question of why UvrY-FLAG crosslinked *in vivo* (though weakly) to the regulatory regions of 286 genes in *E. coli* and 301 genes in *Salmonella* (S2 Table), but exerted little or no regulatory effect in the examples that were tested. To explain these observations, we propose three possible hypotheses: (i) Perhaps other unknown activator(s) are required along with UvrY/SirA for the expression of these genes, but were not available or functioning under the chosen growth conditions. Thus, UvrY/SirA might activate some of these genes under other growth conditions. (ii) Perhaps one or more repressors overrides the influence of UvrY/SirA on the expression of these genes under the conditions tested [62]. To address these two hypotheses, we searched for factors known to regulate the expression of the putative UvrY target genes. This analysis revealed several DNA binding proteins including CRP, FNR and IHF appear to regulate the greatest number of potential UvrY targets (S7 Fig). Hence, it is conceivable that such factors may mask the effects of UvrY on the expression of these genes under our growth conditions. (iii) Because the putative UvrY targets from the ChIP lack the 18nt-long IR sequence found in the regulatory regions of *csrB/C*, it is also possible that some of these genes have degenerate UvrY/SirA binding sites with no functional relevance or perhaps serving as nonspecific holding sites for UvrY/SirA [62, 63].

UvrY expression is activated by CsrA via a putative posttranscriptional mechanism that does not involve the RNA helicase DeaD, a known regulator of *uvrY* translation

In addition to BarA-UvrY, the transcription of CsrB/C is also strongly activated by CsrA [13, 16, 35] by mechanisms that are yet to be determined. Recent evidence suggested that CsrA positively affects *uvrY* expression both at the transcriptional and translational levels [35]. Moreover, CsrA is required for switching BarA from a protein possessing phosphatase activity to kinase activity on UvrY [35]. Thus, we tested the effect of CsrA on the expression, stability and phosphorylation of UvrY-FLAG. The results showed that while CsrA does not affect UvrY-FLAG protein stability (S8A Fig), it has a strong positive effect on UvrY-FLAG expression
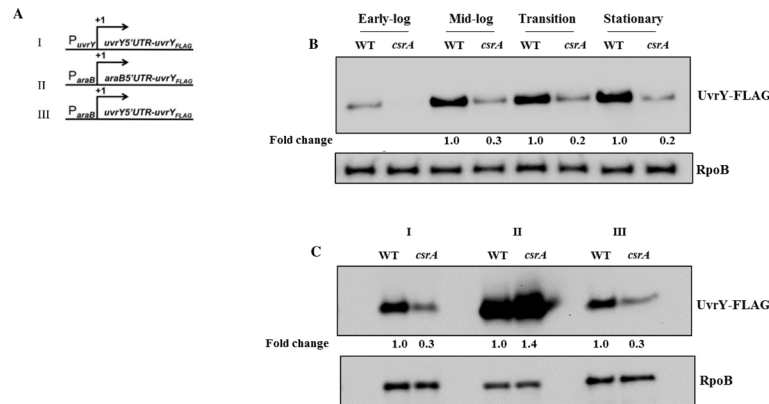
Fig 9. Effect of CsrA on *uvrY* expression. Gene fusions used in this study: (i) native *uvrY* regulatory region and coding sequence containing a FLAG® tag; (ii) *araB* promoter and mRNA nocoding leader fused to the *uvrY-FLAG* coding sequence and (iii) *araB* promoter fused at the transcription start site to the *uvrY* noncoding leader and *uvrY-FLAG* coding sequence (A) as previously depicted (4). Effects of *csrA* disruption on expression of the native *uvrY-FLAG* construct (panel A, i) in LB at different growth phases (B). Effects of *csrA* disruption on expression of the *uvrY-FLAG* constructs shown in panel A (C). Strains were grown in the presence of arabinose, and UvrY-FLAG was detected at mid-log phase of growth by Western blotting. RpoB served as a loading control for these experiments.

doi:10.1371/journal.pone.0145035.g009

(Fig 9B). We observed a 5-fold reduction in UvrY-FLAG levels in a *csrA*::*kan* mutant strain compared to the isogenic wild type strain. These results confirm that CsrA regulates the expression of CsrB and CsrC by activating UvrY expression. Moreover, *in silico* analyses did not reveal any typical CsrA binding sites in the 5'-UTR of the *uvrY* transcript (data not shown) and *uvrY* mRNA was not among the 721 different transcripts that co-purified with CsrA [2], suggesting that the CsrA effect on *uvrY* expression is likely to be indirect.

To assess the genetic effects of CsrA on UvrY expression, we analyzed fusion derivatives that replaced the *uvrY* promoter or the promoter and noncoding leader with those from *araB* (Fig 9A) [4]. Results from these experiments showed that regulation by CsrA requires the *uvrY* noncoding leader but not the promoter region (Fig 9C). To test whether CsrA requires sequence from the *uvrY* coding region for regulation, we measured the expression of a *uvrY'-'lacZ* translational reporter fusion containing the leader and 12 or 22 codons of *uvrY* in wild type and *csrA*::*kan* strains (Fig 10A). The two different leaders were chosen because previous studies showed that the DeaD RNA helicase is needed to override inhibitory long-distance mRNA base-pairing between the *uvrY* noncoding leader and the proximal *uvrY* coding segment, and that DeaD regulates *uvrY* expression from gene fusions with 22 or more *uvrY* codons, but not from fusions with less than this, e.g. 12 codons, which do not permit the inhibitory base-pairing [4]. Our results showed that expression of both fusions (22 and 12 codon) was decreased in the *csrA*::*kan* strain by 2.8 and 2.6-fold, respectively (Fig 10B), suggesting that the effects of CsrA are not mediated via DeaD. We also tested the effect of CsrA on the expression of a DeaD-FLAG fusion, expressed from the native chromosomal locus, and confirmed that CsrA does not regulate *deaD* expression (Fig 10C).

Furthermore, another RNA helicase, SrmB, was shown to regulate the transcription of CsrB and CsrC RNAs by an undefined mechanism, which is different than that of DeaD [4]. We, therefore, examined the effect of CsrA on the levels of SrmB-FLAG and observed no effect on the expression of this RNA helicase (Fig 10D). Thus, CsrA does not appear to regulate expression of either of the DeaD-box RNA helicases that activate CsrB and CsrC transcription.

Finally, in a screen to find novel direct CsrA targets by RNAseq analysis, *ihfA* mRNA was found to co-purify with CsrA [2], suggesting that CsrA might regulate production of the IHF
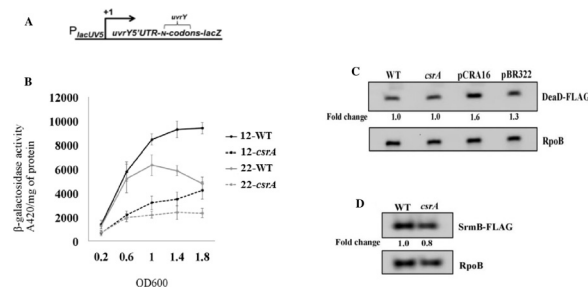
Fig 10. CsrA activates *uvrY* expression without affecting DeaD or SrmB RNA helicase levels. The *uvrY* gene fusions used in this study were previously depicted in (4): *lacUV5* promoter fused at the transcription start site to the *uvrY* mRNA leader and N (12 or 22) *uvrY* codons fused in frame to *lacZ* (A). The effect of *csrA* disruption on expression of P*lacUV5-uvrY'–'lacZ* reporter fusions is shown (B). Cells were grown in LB and harvested at various times throughout growth and assayed for β-galactosidase specific activity (A$_{420}$/mg protein). The values represent the average of two independent experiments. Error bars depict standard error of the means. Western blots showing effects of *csrA* on the level of DeaD-FLAG (C) or SrmB-FLAG (D) in an MG1655 derivative that expresses the corresponding FLAG-tagged gene from its native genomic locus are shown. Effect of *csrA::kan* complemented by a *csrA* expression plasmid (pCRA16) or with a control plasmid (pBR322) is also shown for DEAD-FLAG. RpoB served as a loading control for these blots.

doi:10.1371/journal.pone.0145035.g010

protein. Therefore, we hypothesized that in addition to activating *uvrY* expression, CsrA might also activate the expression of *csrB* by activating the expression of IHF. However, we observed no effect of CsrA on IhfA-FLAG or IhfB-FLAG levels when these fusions were expressed from their native chromosomal loci (S8B Fig). Altogether, our results indicate that CsrA activates *csrB/C* transcription by activating *uvrY* expression, most likely through a posttranscriptional mechanism that is yet to be defined.

Effects of ppGpp and DksA, mediators of the stringent response, on *csrB* expression

Previously, ppGpp and DksA were shown to strongly activate expression of CsrB and CsrC in *E. coli* [2]. However, the mechanism for these effects is yet to be elucidated. In an attempt to define this mechanism, we hypothesized that ppGpp and DksA might activate the expression of factor(s) that are known to activate CsrB/C transcription i.e., UvrY, IHF, CsrA, DeaD and/or SrmB. Previous studies showed that ppGpp and DksA positively affect CsrA levels, but this effect appears to be too modest to account for their strong effects on CsrB/C levels [2]. Therefore, we first determined whether DksA and ppGpp affect the *in vivo* levels of UvrY-FLAG protein (Western blot) and the *in vivo csrB* DNA binding (ChIP-PCR) by UvrY in Δ*dksA*, Δ*relA* (encoding the major ppGpp synthase) and isogenic wild type strains. It was found that ppGpp and DksA had weak or negligible effects on UvrY levels (Fig 11), *in vivo* binding of UvrY to *csrB* (Fig 12), and the *in vivo* phosphorylation status of the UvrY protein (S3 Fig). Next, we tested the effects of DksA and ppGpp on IHF expression. The results showed no effect of DksA or ppGpp on IhfA-FLAG (Fig 13A) or IhfB-FLAG levels under our growth conditions (Fig 13B).

Because ppGpp and DksA did not substantially affect the *in vivo* expression of IHF or alter the *in vivo* binding of UvrY-P to *csrB* DNA, we decided to test their direct effects on the *in vitro* transcription of *csrB* and *csrC* genes. However, neither gene was expressed in defined transcription reactions that contained UvrY-P and IHF in the case of *csrB* and basal reaction components (5nM RNAP; 40 mM Tris·HCl, pH 7.9; NaCl (165 mM); 5% glycerol; 10 mM MgCl$_2$; 1 mM DTT; 0.1 µg/µl BSA; 500 µMATP; 200 µM CTP and UTP; 10 µMGTP and [α-32P] GTP (2.5 µCi). This suggests that unknown factors may be required for *csrB/C*
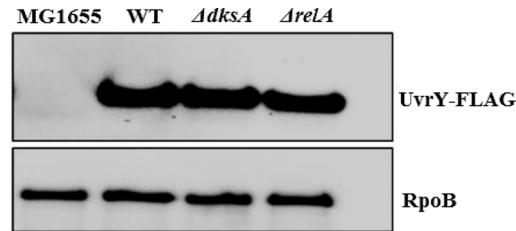
Fig 11. Effect of DksA and RelA on UvrY-FLAG levels. Western blotting of UvrY-FLAG levels in strains MG1655 (no FLAG fusion), WT (MG1655 expressing UvrY-FLAG from the *uvrY* genomic locus), and isogenic ΔdksA and ΔrelA strains. Proteins were collected from cultures grown in LB medium to mid-exponential growth phase (~OD$_{600}$ of 0.6). RpoB served as a loading control.

doi:10.1371/journal.pone.0145035.g011

transcription (data not shown). We next used coupled transcription-translation in S-30 extracts to determine whether the addition of ppGpp and DksA would directly regulate expression of *csrB/C* in the presence of other cellular factors present in the S-30 extracts but not the defined transcription reactions. As previously reported [13, 16], UvrY-P strongly stimulated the expression of the full-length β-galactosidase from *csrB-lacZ* (Fig 14) and *csrC-lacZ* (S9 Fig) transcriptional fusions in this assay. A series of truncated β-galactosidase products was also observed in these reactions, which quantitatively responded to activators similarly to the full-length protein. While expression in the absence of UvrY-P was weak, the addition of ppGpp or ppGpp and DksA to the UvrY-deficient *csrB-lacZ* reaction caused a modest increase in expression (Fig 14). In the presence of UvrY-P, addition of ppGpp alone modestly activated *csrB-lacZ* expression (1.5-fold). DksA alone also had weak effects in the presence of UvrY-P (1.4-fold), while the addition of both ppGpp and DksA led to a 1.7-fold increase in *csrB-lacZ* expression. Addition of ppGpp alone, in the absence of UvrY-P, resulted in a slight increase in *csrC-lacZ* expression (S9 Fig). However, ppGpp and/or DksA failed to activate *csrC-lacZ* expression in reactions containing UvrY-P. We conclude that ppGpp and DksA directly activate *csrB* expression, at least partially accounting for the stimulatory effect of these regulators on the *in vivo* expression of this gene [2]. However, ppGpp and/or DksA may be indirectly involved in or require a factor that was deficient in our assays for *csrC* expression.

As mentioned above, DeaD and SrmB activate *csrB* and *csrC* transcription by distinct mechanisms [4]. While DeaD activates *uvrY* translation, the mechanism by which SrmB activates *csrB/C* transcription is still not defined. Our ChIP-PCR data revealed that SrmB is required for normal binding of UvrY to *csrB in vivo* (S2 Fig). This effect of SrmB on UvrY binding occurs without it altering UvrY or UvrY-P levels [4]. Because neither ppGpp nor DksA had substantial
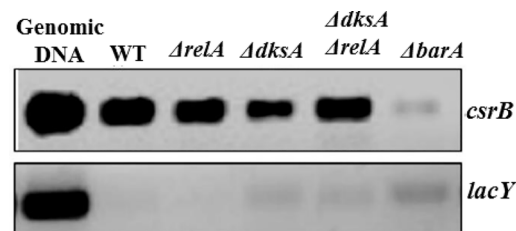

Fig 12. Effect of DksA and RelA on *in vivo* binding of UvrY to *csrB* promoter. The effect of DksA and RelA on *in vivo* binding of UvrY to *csrB* promoter was determined by ChIP-PCR assay in a WT (MG1655 expressing UvrY-FLAG) and isogenic ΔdksA, ΔrelA and ΔbarA strains. Agarose gel showing PCR amplification of *csrB* promoter region recovered from each strain. The *lacY* gene served as a negative control in this experiment.
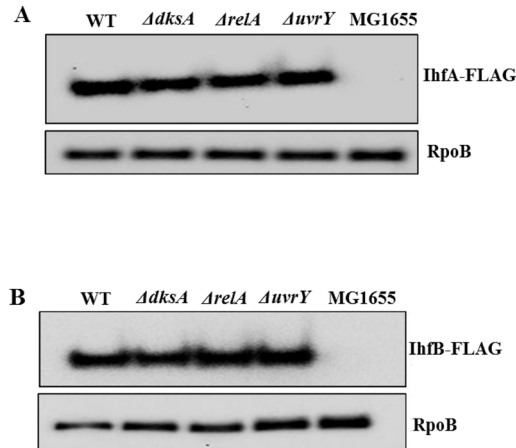
doi:10.1371/journal.pone.0145035.g012

Fig 13. Effect of DksA, RelA and UvrY on the expression of IHF subunits. Western blotting of IhfA-FLAG (A) and IhfB-FLAG (B) proteins examined in WT (MG1655 expressing *ihfB-FLAG* or *ihfA-FLAG* fusions from the native genomic loci) and an MG1655 control lacking the *FLAG* fusions. Proteins from isogenic strains with Δ*dksA*, Δ*relA* and Δ*uvrY* disruption are as shown. Proteins were collected from cultures grown in LB medium to mid-exponential growth phase (~OD$_{600}$ of 0.6). RpoB loading controls for these analyses are also shown.

doi:10.1371/journal.pone.0145035.g013

effects on *in vivo* binding of UvrY to *csrB* DNA (Fig 12) we infer that the mechanism by which ppGpp and DksA activate *csrB/C* transcription does not involve SrmB and vice versa.

In silico analysis of other possible targets of UvrY (SirA) binding

In many cases, transcription factors compete for binding to their DNA binding sites with other transcription factors, which may play antagonistic roles in the regulation of the target gene(s) [64]. Hence, we reasoned that there could be additional targets of UvrY/SirA in the genomes of *E. coli* / *Salmonella* that were not captured by ChIP-exo. To test this hypothesis, we performed
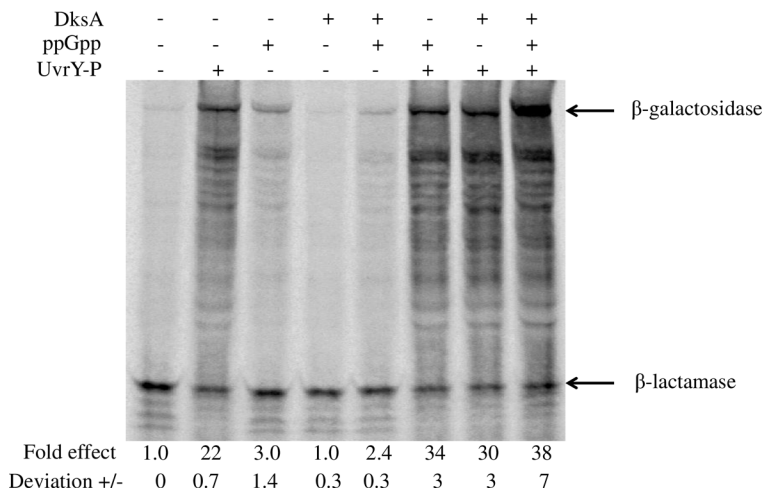

DksA	-	-	-	+	+	-	+	+
ppGpp	-	-	+	-	+	+	-	+
UvrY-P	-	+	-	-	-	+	+	+
Fold effect	1.0	22	3.0	1.0	2.4	34	30	38
Deviation +/-	0	0.7	1.4	0.3	0.3	3	3	7

Fig 14. *In vitro* transcription-translation of a supercoiled plasmid-encoded *csrB-lacZ* transcriptional fusion. Reactions contained pLFXcsrB-lacZ plasmid (2 µg), UvrY-P (2.3 µM), ppGpp (250 µM) and/or DksA (2 µM) as indicated. Incorporation of ^{35}S-labeled methionine into protein products was detected by SDS PAGE followed by phosphorimaging. Signal intensity of the full length protein was determined using Quantity One software. The fold-effects of regulatory factors were determined with respect to the control reaction lacking the factors, after normalization against the internal control, β-lactamase, which was encoded on the same plasmid. Absolute deviation for each reaction was determined from two independent experiments.

doi:10.1371/journal.pone.0145035.g014

in silico analysis using the *Ab Initio* Motif Identification Environment (AIMIE) database [65]. We scanned the *E. coli* genome using the first six bases of the 18 bp (TGTAAGNNNNNNNCT-TACA) UvrY binding sequence, followed by manually checking the presence of the rest of the IR DNA sequence in the regulatory region of each discovered putative target. In this way, we identified 19 putative target genes containing the 18-bp UvrY-P IR sequence either perfectly or near perfectly conserved in their regulatory regions (S10 Fig). Next, we searched for factors known to directly regulate the expression of these putative targets and compared the respective DNA binding motif of each factor with the 18 bp IR UvrY binding motif. Out of the 19 putative targets, regulatory factors were previously established for 5 of them (S11 Fig). When we compared the respective DNA binding motifs of each factor with the IR binding motif of UvrY, we found an overlap in all of them (S11 Fig), supporting the possibility that undiscovered direct target sequences for UvrY binding might have gone undetected in our experiments. Whether these putative target sequences function in UvrY regulation under other growth conditions will require future investigation.

Phylogenetic distribution of BarA, UvrY, CsrA and FliW

The BarA-UvrY TCS exerts global effects on gene expression by activating CsrB and CsrC transcription, thus controlling CsrA activity [16, 35]. This signaling pathway is common to the commensal bacterium *E. coli*, the pathogen *Salmonella*, as well as a variety of other γ-proteobacterial pathogens [66]. In contrast, little is known about the workings of the Csr system in other species that possess *csrA* homologs, but have not been shown to express CsrA-inhibitory sRNAs. Recently, Mukherjee et al. have shown that the FliW protein of *B. subtilis* binds to CsrA and antagonizes its activity, thus preventing it from binding to the flagellin mRNA, *hag* [41, 67]. Because BarA-UvrY is devoted to transcription of CsrB/C sRNAs in *E. coli* and *Salmonella* and *fliW* is absent in species known to produce CsrA-inhibitory sRNAs [41], we hypothesized that BarA-UvrY and FliW might represent different modules for regulating CsrA activity in different bacterial species. To test this hypothesis, we examined the phylogenetic distributions of CsrA, BarA-UvrY and FliW across fully-sequenced bacteria (S3 Table). We found that, for species encoding at least one readily identifiable CsrA ortholog, the presence of BarA-UvrY and FliW were strongly anti-correlated at species level (Spearman correlation = -0.97, p = 1.10×10^{-205}) and genus level (Spearman correlation = -0.95, p = 7.31×10^{-74}) (Fig 15). Of the 346 genomes encoding CsrA, 340 also encoded either BarA-UvrY or FliW, but not both, and only 6 species might encode both BarA-UvrY and FliW systems: *Desulfosporosinus acidiphilus*, *D. meridiei*, *D. orientis*, *D. baculatum*, *Magnetococcus marinus*, and *Candidatus Sulfuricurvum sp* (Fig 15 and S3 Table). Interestingly, in all cases where both BarA/UvrY and FliW may function, only the BarA component appears to be present, which may indicate that this two-component system has lost or is losing its function in these species. Additional experimental data are needed in order to confirm this hypothesis.

Accurate and complete identification of orthologs across a broad taxonomic range is a challenging problem, with potential for both false-positives and false-negatives [68]. Although difficult to detect, any strong biases in the methodology we used to identify orthologous genes could affect our results. To address this concern, we used two additional and widely-used methods to identify alternative ortholog sets: KEGG orthology database (KO) [53], and UniProt reference clusters of orthologs (UniRef) [54]. We found that, regardless of the method used to identify orthologs, FliW and BarA-UvrY were strongly negatively correlated, particularly when conditioning on the presence of CsrA (S3 Table). In addition, we applied sequence similarity filtering and manual curation to reduce false positives and false negatives in the NCBI orthologs data set.
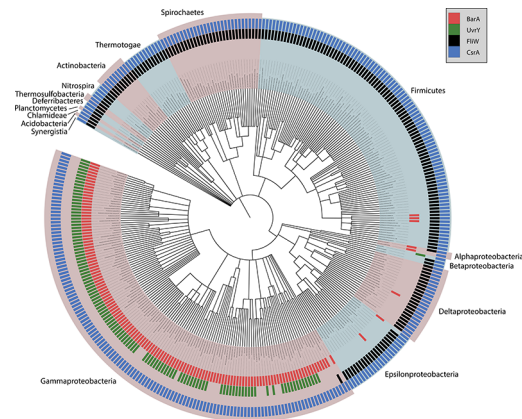
Fig 15. The distribution of CsrA, BarA, UvrY and FliW across phylogenies indicates a strong anti-correlation between BarA/UvrY and FliW. Depicted is a reference phylogeny of fully-sequenced genomes encoding CsrA and at least one of BarA, UvrY or FliW, plotting their presence/absence data obtained from orthology databases. The results show that the presence of BarA-UvrY and FliW are significantly anti-correlated.

doi:10.1371/journal.pone.0145035.g015

After the sequence similarity filtering and manual curation, we found several false positives for CsrA in the Streptococcus genus, such as misannotated heavy metal stress response proteins (NCBI identification numbers (GI): 15675047, 13622200, 21904470), peptide methionine sulfoxide reductases (GI: 222114035, 134272076, 209540564, 24638057, 342165139, 342165138), and putative CsrA homologs that lack the N-terminal and C-terminal conserved regions of CsrA (GI: 895760047, 882844105, 882819224). Most of the main false negatives that we identified are BarA and UvrY from genomes that match the representative homologs from *E. coli* within the similarity and coverage thresholds (see methods) such as *Pseudomonas* (NCBI genome accession numbers: NC_022594, NC_022591, NC_022361, NC_022360), *Xanthomonas* (NC_020815, NC_017271, NC_017267, NC_016010, NC_013722), *Pseudoxanthomonas* (NC_014924), *Shewanella* (NC_009052, NC_009665, NC_008321, NC_008322, NC_017566, NC_016901), and other members of the γ-Proteobacteria class ([S3 Table](#)).

These false positives and false negatives caused a small increase of 2% in the negative correlation, reinforcing the present conclusions (Spearman correlation = -0.95 vs. -0.97, comparing the unfiltered and filtered data sets, respectively) ([S3 Table](#)). These results argue against methodological bias as strongly affecting our results. Together, these results indicate that the negative correlation between FliW and BarA-UvrY regulatory systems is highly unlikely to be artifactual and thus represents a biologically relevant observation. The implications of this observation for the possible regulation of CsrA activity by inhibitory sRNAs in FliW-encoding species will require additional investigation to unravel.

Conclusions

Using ChIP-exo [34], we probed the complete repertoire of UvrY (SirA) DNA binding sites in the genomes of *E. coli* and *Salmonella*. We discovered that the *csrB*/C genes are by far the strongest direct targets of UvrY in these species. UvrY binds specifically to an 18 nt palindromic sequence in the promoter regions of *csrB/C* and exhibited an almost absolute requirement for phosphorylation by BarA for this binding *in vivo* under growth conditions examined. UvrY-P requires IHF for optimal binding to and activation of *csrB* but not *csrC*. CsrA activates *csrB/C* transcription by activating *uvrY* expression by an undefined mechanism, which may require the noncoding mRNA leader, but does not involve the other known posttranscriptional

regulator of *uvrY*, the DeaD-box RNA helicase DeaD [4]. The RNA DEAD-box helicase SrmB, which also activates *csrB/C* transcription in *E. coli* [4], promoted binding of UvrY to *csrB* DNA *in vivo* without affecting the expression of other factors known to activate *csrB* expression. This suggests that SrmB may regulate unknown factor(s) involved in *csrB* transcription. The stringent response factors ppGpp and DksA activate CsrB/C expression *in vivo* [2] and were found to modestly activate *csrB* expression *in vitro* in S-30 extracts. Whether this involves direct binding to RNA polymerase remains to be determined.

Genomic loci that crosslinked weakly to UvrY were identified proximal to the promoter regions of 286 genes in *E. coli* and 301 genes in *Salmonella*, respectively (S2 Table). However, further analysis showed weak or negligible regulatory effects of UvrY on the expression of the genes that were tested (S6A–S6C Fig). Hence, we conclude that, under the growth conditions that we have examined, UvrY-P exerts its global effects on gene expression almost entirely by activating the transcription of CsrB and CsrC. We suspect that most of the genes that have been found to respond to BarA and UvrY in *E. coli* [31] and their orthologs in other species [30, 32] are indirect targets of UvrY, which are regulated by CsrA.

The BarA-UvrY/Csr signaling pathway has been studied in the commensal bacterium *E. coli*, the pathogen *Salmonella*, and a number of other γ-proteobacterial pathogens [66]. *B. subtilis*, the only Gram-positive bacterium in which CsrA has been studied to date, uses FliW as an α-CsrA protein, which binds to and inhibits CsrA activity [41, 67]. While this is only the first such example, FliW is present in diverse species, where it may act as an inhibitor of CsrA. For species encoding at least one readily identifiable CsrA ortholog, the presence of BarA-UvrY and FliW were strongly anti-correlated. This suggests that while γ-Proteobacteria use the BarA-UvrY TCS to control CsrA by activating the transcription of its sRNA antagonists, members of the β-Proteobacteria, δ-Proteobacteria, ε-Proteobacteria, Firmicutes, Spirochaetes, Thermotogae, Actinobacteria, Nitrospira, Thermosulfobacteria, Deferribacteres, Planctomycetes, Chlamideae, Acidobacteria, and Synergistia may use FliW to regulate CsrA activity.

Supporting Information

S1 Fig. SirA-FLAG and UvrY-His$_6$ are functional *in vivo*. Western blot showing expression of SirA-FLAG (A) and Northern blots showing CsrB levels in 14028S (wild type *Salmonella*), 14028S strain with *sirA-FLAG* fusion integrated at the native *sirA* locus and *sirA* and *csrB* deletion strains (B). CsrB levels in MG1655, *uvrY* deletion and UvrY-His$_6$ (expressed from pET24-a expression vector) *E. coli* strains (C). For Western blotting, RpoB loading controls shown (A). For Northern blotting, the 16S/23S rRNA loading controls are shown. Cultures were grown in LB to mid-exponential growth phase (OD$_{600}$ of 0.6).
(TIFF)

S2 Fig. ChIP specificity confirmation by PCR. Polymerase chain reaction was used to confirm the specificity of ChIP assay. Primers (S1 Table) annealing to the promoter regions of *csrB*, *lacY* and 16S rDNA (*rrsH*) were used to amplify the promoters of *csrB*, *lacY* and/or 16S rDNA genes from DNA that was crosslinked and immunoprecipitated from *E. coli* (panel A) or *Salmonella* (panel B). In these analysis, *csrB* was used as a positive control and *lacY* and 16S rDNA (*rrsH*) were used as negative controls for *E. coli* and *Salmonella*, respectively.
(TIFF)

S3 Fig. Effect of BarA, IHF, DksA, RelA and CsrA on UvrY-FLAG and UvrY-FLAG-P levels. Phos-tag SDS-PAGE with Western blotting was used for detection of the phosphorylated (P-UvrY-FLAG) and non-phosphorylated (UvrY-FLAG) protein levels expressed in a WT (MG1655 expressing UvrY-FLAG) and isogenic Δ*barA*, Δ*dksA*, Δ*relA*, Δ*ihfA*, Δ*ihfB*, *csrA::kan*

and Δ*uvrY* strain. Cultures were grown in LB to mid-exponential growth phase (OD$_{600}$ of 0.6). The relative levels and % of phosphorylation of UvrY in the WT, *barA*, *dksA*, *relA*, *ihfA*, *ihfB* and *csrA* are: 1.0, 1.2, 1.0, 1.0, 1.2, 1.2, 0.23 (UvrY levels) and 7%, 1%, 6.8%, 6.9%, 9%, 9% and 8% (% of UvrY phosphorylation), respectively.
(TIFF)

S4 Fig. DNase I footprinting of *csrC* DNA by phosphorylated and non-phosphorylated UvrY. A ^{32}P-end labeled DNA probe (reverse strand) that included both the upstream and downstream putative UvrY binding sites was used ([Fig 1E](#)). Reactions in all lanes except lanes 1 contained DNase I (0.025U/12.5ul reaction). Reactions in lanes 3–6 and lanes 7–10 contained 0.25, 0.35, 0.5, 0.7 µM of phosphorylated and non-phosphorylated UvrY-His$_6$, respectively. Lane 2 reaction contained no UvrY.
(TIFF)

S5 Fig. Effect of IHF on UvrY-FLAG levels. Western blotting of UvrY-FLAG levels in strains MG1655 (no FLAG fusion), WT (MG1655 expressing UvrY-FLAG), and isogenic Δ*ihfA* and Δ*ihfB* strains. Cultures were grown in LB to mid-exponential growth phase (OD$_{600}$ of 0.6). RpoB loading control is also shown.
(TIFF)

S6 Fig. Examination of putative UvrY targets discovered by ChIP-exo. Northern blot showing effect of UvrY and Csr factors on SpoT42 sRNA (A). Cultures of MG1655 and the isogenic mutants indicated were grown in Kornberg medium containing 0.5% glucose, to early stationary growth phase (OD$_{600}$ of 2.0). The 16S/23S rRNA loading controls are also shown. Western blot showing effect of *uvrY* deletion on FhuF-FLAG protein (B). Cultures were grown in LB to mid-exponential growth phase (OD$_{600}$ of 0.6), at which point dipyridyl was added to culture (1mM final concentration). Samples were collected before and 10 min after the addition of dipyridyl. RpoB loading control is also shown. Electrophoretic gel mobility shift assay showing UvrY binding to *cspA*, *spf*, *fhuF* and *csrB* DNA (C). Phosphorylated (UvrY-P) UvrY-His$_6$ binding to *spf*, *cspA*, *fhuF*, *csrB* and *rrlE* DNA was tested by EMSA as shown. The *spf*, *cspA*, *fhuF* and *csrB* DNA probes used in this experiments encompass the ChIP-exo derived putative UvrY binding sites discovered in the promoter region of each gene (shown in [Fig 1](#) and [S2 Table](#)). The DNA probes (0.5 nM) were incubated at room temperature with increasing concentration of *in vitro* phosphorylated UvrY-His$_6$ protein. End-labeled 0.5 nM *rrlE* and 50-fold cold *csrB* (for the specific competitor, marked with *, 0.65 µM UvrY-P was used) were also used as non-specific and specific competitors, respectively. The DNA-protein complexes were resolved in a non-denaturing 7% polyacrylamide gel. Shifted protein-DNA complex is indicated in black arrows. Effect of UvrY on the expression of putative sRNA targets (D). Northern blots showing effect of UvrY on the expression of several sRNA genes. Cultures were grown in Kornberg, supplemented with 0.5% glucose, to mid-exponential (OD$_{600}$ of 0.6), transition to stationary (OD$_{600}$ of 1.2) and stationary growth phases (OD$_{600}$ of 3.0). The 5S rRNA loading control is also shown.
(TIF)

S7 Fig. List of transcription factors known to regulate expression of the ChIP-exo derived putative UvrY targets (listed in [S2 Table](#)).
(TIFF)

S8 Fig. Effect of CsrA on UvrY-FLAG protein stability and expression of IHF subunits. Western blot UvrY-FLAG protein stability (A) in MG1655 (no FLAG fusion), WT (MG1655 with a *uvrY-FLAG* fusion integrated at the *uvrY* locus) and isogenic *csrA* mutant. Cells were

grown in LB to mid-exponential growth phase (OD$_{600}$ of 0.6) at which point tetracycline and chloramphenicol were added and cultures were sampled thereafter at the times shown. Western blotting of IhfA-FLAG and IhfB-FLAG proteins (B) examined in MG1655 or an isogenic *csrA*::*kan* mutant expressing *ihfA-FLAG* or *ihfB-FLAG* fusions from the native genomic loci (WT). Cultures were grown in LB to mid-exponential growth phase (~OD$_{600}$ of 0.6). RpoB loading controls for these analyses are also shown.
(TIFF)

S9 Fig. *In vitro* transcription-translation of a supercoiled plasmid-encoded *csrC-lacZ* fusion. Reactions contained pLFXcsrC-lacZ (4 μg), UvrY-P (2.3 μM), ppGpp (250 μM) and/or DksA (2 μM) as indicated. Incorporation of ^{35}S-labeled methionine into protein products was detected by SDS PAGE with phosphorimaging. Signal intensity was determined using Quantity One software. The fold-effects of regulatory factors were determined with respect to the control reaction lacking these factors, after normalization against β-lactamase as an internal control. Absolute deviation was determined from two independent experiments.
(TIFF)

S10 Fig. Putative UvrY targets derived by *in silico* analysis. 19 Putative UvrY targets were derived by *in silico* analysis using the *Ab Initio* Motif Identification Environment (AIMIE) database (46). This was done by scanning the *E. coli* genome using the first six bases of the 18bp-long (TGTAAGNNNNNNCTTACA) UvrY consensus binding sequence, followed by manually checking the presence of the rest of the IR DNA sequence in the regulatory region of each discovered putative target. The predicted UvrY binding sequence in the promoter region of each putative target in comparison to the UvrY consensus sequence is shown. Distance from the center of the predicted UvrY binding sequence to the known transcription start site (TSS) of each putative target is shown. Nucleotides marked in red are the mismatch between the UvrY consensus sequence and the predicted UvrY binding sequence.
(TIFF)

S11 Fig. Overlap between predicted UvrY binding site and consensus DNA binding site of known regulators of putative UvrY targets derived by *in silico* analysis. Shown in this figure are five of the 19 putative UvrY targets identified by *in silico* analysis (S9 Fig), for which regulatory factors were previously established. The list of the five putative targets, their respective known regulators, the consensus DNA binding site of each regulator (capitalized) and the overlap between the consensus binding site of the known regulator and the predicted UvrY binding site within the promoter of each putative target (underlined) is shown.
(TIFF)

S1 Table. List of strains, plasmids, bacteriophages and primers used in this study
(DOCX)

S2 Table. List of putative UvrY-SirA targets identified by ChIP-exo.
(XLSX)

S3 Table. Distribution of BarA, UvrY, CsrA, and FliW at species and genus level. We identified CsrA, BarA, UvrY and FliW orthologs from all fully-sequenced bacterial genomes in the NCBI genomes database (http://www.ncbi.nlm.nih.gov/genome/). The results were compared to other orthology databases and filtered by sequence similarity and alignment coverage in order to minimize the number of false positives/negatives (see Methods). At both species and genus levels, the results exhibited a strong anti-correlation between BarA-UvrY and FliW.
(XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: TRZ CAV YL AP AHP RD DT BK DG BMMA TR. Performed the experiments: TRZ CAV YL AP AHP RD DT. Analyzed the data: TRZ CAV YL AP AHP RD BK BMMA TR. Contributed reagents/materials/analysis tools: BK BMMA TR. Wrote the paper: TRZ CAV AP AHP RD BK TR.

References

1. Wei BL, Brun-Zinkernagel AM, Simecka JW, Prüss BM, Babitzke P, Romeo T. Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. Mol Microbiol. 2001; 40(1):245–56. PMID: 11298291

2. Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, et al. Circuitry linking the Csr and stringent response global regulatory systems. Mol Microbiol. 2011; 80(6):1561–80. doi: 10.1111/j.1365-2958.2011.07663.x PMID: 21488981

3. Romeo T, Vakulskas CA, Babitzke P. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. Environ Microbiol. 2013; 15(2):313–24. doi: 10.1111/j.1462-2920.2012.02794.x PMID: 22672726

4. Vakulskas CA, Pannuri A, Cortés-Selva D, Zere TR, Ahmer BM, Babitzke P, et al. Global effects of the DEAD-box RNA helicase DeaD (CsdA) on gene expression over a broad range of temperatures. Mol Microbiol. 2014; 92(5):945–58. doi: 10.1111/mmi.12606 PMID: 24708042

5. Pannuri A, Yakhnin H, Vakulskas CA, Edwards AN, Babitzke P, Romeo T. Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. J Bacteriol. 2012; 194(1):79–89. doi: 10.1128/JB.06209-11 PMID: 22037401

6. Yakhnin AV, Baker CS, Vakulskas CA, Yakhnin H, Berezin I, Romeo T, et al. CsrA activates *flhDC* expression by protecting flhDC mRNA from RNase E-mediated cleavage. Mol Microbiol. 2013; 87 (4):851–66. doi: 10.1111/mmi.12136 PMID: 23305111

7. Yakhnin H, Baker CS, Berezin I, Evangelista MA, Rassin A, Romeo T, et al. CsrA represses translation of *sdiA*, which encodes the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of *sdiA* mRNA. J Bacteriol. 2011; 193(22):6162–70. doi: 10.1128/JB.05975-11 PMID: 21908661

8. Yakhnin H, Yakhnin AV, Baker CS, Sineva E, Berezin I, Romeo T, et al. Complex regulation of the global regulatory gene *csrA*: CsrA-mediated translational repression, transcription from five promoters by Eσ⁷⁰ and Eσ(S), and indirect transcriptional activation by CsrA. Mol Microbiol. 2011; 81(3):689–704. doi: 10.1111/j.1365-2958.2011.07723.x PMID: 21696456

9. Mercante J, Suzuki K, Cheng X, Babitzke P, Romeo T. Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance. J Biol Chem. 2006; 281(42):31832–42. PMID: 16923806

10. Mercante J, Edwards AN, Dubey AK, Babitzke P, Romeo T. Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. J Mol Biol. 2009; 392(2):511–28. doi: 10.1016/j.jmb.2009.07.034 PMID: 19619561

11. Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM. Identification and molecular characterization of *csrA*, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. J Bacteriol. 1993; 175(15):4744–55. PMID: 8393005

12. Liu MY, Gui G, Wei B, Preston JF, Oakford L, Yüksel U, et al. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. J Biol Chem. 1997; 272 (28):17502–10. PMID: 9211896

13. Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, et al. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. Mol Microbiol. 2003; 48(3):657–70. PMID: 12694612

14. Altier C, Suyemoto M, Ruiz AI, Burnham KD, Maurer R. Characterization of two novel regulatory genes affecting *Salmonella* invasion gene expression. Mol Microbiol. 2000; 35(3):635–46. PMID: 10672185

15. Suzuki K, Babitzke P, Kushner SR, Romeo T. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes Dev. 2006; 20 (18):2605–17. PMID: 16980588

16. Suzuki K, Wang X, Weilbacher T, Pernestig AK, Melefors O, Georgellis D, et al. Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*. J Bacteriol. 2002; 184(18):5130–40. PMID: 12193630

17. Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol. 2007; 10(2):156–63. PMID: 17383221

18. Martínez LC, Martínez-Flores I, Salgado H, Fernández-Mora M, Medina-Rivera A, Puente JL, et al. In silico identification and experimental characterization of regulatory elements controlling the expression of the *Salmonella csrB* and *csrC* genes. J Bacteriol. 2014; 196(2):325–36. doi: 10.1128/JB.00806-13 PMID: 24187088

19. Teplitski M, Goodier RI, Ahmer BM. Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*. J Bacteriol. 2003; 185(24):7257–65. PMID: 14645287

20. Pernestig AK, Melefors O, Georgellis D. Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*. J Biol Chem. 2001; 276(1):225–31. PMID: 11022030

21. Sahu SN, Acharya S, Tuminaro H, Patel I, Dudley K, LeClerc JE, et al. The bacterial adaptive response gene, *barA*, encodes a novel conserved histidine kinase regulatory switch for adaptation and modulation of metabolism in *Escherichia coli*. Mol Cell Biochem. 2003; 253(1–2):167–77. PMID: 14619967

22. Chavez RG, Alvarez AF, Romeo T, Georgellis D. The physiological stimulus for the BarA sensor kinase. J Bacteriol. 2010; 192(7):2009–12. doi: 10.1128/JB.01685-09 PMID: 20118252

23. Huang Y, Suyemoto M, Garner CD, Cicconi KM, Altier C. Formate acts as a diffusible signal to induce *Salmonella* invasion. J Bacteriol. 2008; 190(12):4233–41. doi: 10.1128/JB.00205-08 PMID: 18424519

24. Lawhon SD, Maurer R, Suyemoto M, Altier C. Intestinal short-chain fatty acids alter *Salmonella typhimurium* invasion gene expression and virulence through BarA/SirA. Mol Microbiol. 2002; 46(5):1451–64. PMID: 12453229

25. Martínez LC, Yakhnin H, Camacho MI, Georgellis D, Babitzke P, Puente JL, et al. Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the *Salmonella* SPI-1 and SPI-2 virulence regulons through HilD. Mol Microbiol. 2011; 80 (6):1637–56. doi: 10.1111/j.1365-2958.2011.07674.x PMID: 21518393

26. Goodier RI, Ahmer BM. SirA orthologs affect both motility and virulence. J Bacteriol. 2001; 183 (7):2249–58. PMID: 11244064

27. Reimmann C, Beyeler M, Latifi A, Winteler H, Foglino M, Lazdunski A, et al. The global activator GacA of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. Mol Microbiol. 1997; 24(2):309–19. PMID: 9159518

28. Hammer BK, Tateda ES, Swanson MS. A two-component regulator induces the transmission phenotype of stationary-phase *Legionella pneumophila*. Mol Microbiol. 2002; 44(1):107–18. PMID: 11967072

29. Wong SM, Carroll PA, Rahme LG, Ausubel FM, Calderwood SB. Modulation of expression of the ToxR regulon in Vibrio cholerae by a member of the two-component family of response regulators. Infect Immun. 1998; 66(12):5854–61. PMID: 9826365

30. Broberg M, Lee GW, Nykyri J, Lee YH, Pirhonen M, Palva ET. The global response regulator ExpA controls virulence gene expression through RsmA-mediated and RsmA-independent pathways in *Pectobacterium wasabiae* SCC3193. Appl Environ Microbiol. 2014; 80(6):1972–84. doi: 10.1128/AEM. 03829-13 PMID: 24441162

31. Oshima T, Aiba H, Masuda Y, Kanaya S, Sugiura M, Wanner BL, et al. Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. Mol Microbiol. 2002; 46(1):281–91. PMID: 12366850

32. Krin E, Derzelle S, Bedard K, Adib-Conquy M, Turlin E, Lenormand P, et al. Regulatory role of UvrY in adaptation of *Photorhabdus luminescens* growth inside the insect. Environ Microbiol. 2008; 10 (5):1118–34. doi: 10.1111/j.1462-2920.2007.01528.x PMID: 18248456

33. Brencic A, McFarland KA, McManus HR, Castang S, Mogno I, Dove SL, et al. The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. Mol Microbiol. 2009; 73(3):434–45. doi: 10.1111/j. 1365-2958.2009.06782.x PMID: 19602144

34. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell. 2011; 147(6):1408–19. doi: 10.1016/j.cell.2011.11.013 PMID: 22153082

35. Camacho MI, Alvarez AF, Gonzalez Chavez R, Romeo T, Merino E, Georgellis D. Effects of the Global Regulator CsrA on the BarA/UvrY Two-Component Signaling System. J Bacteriol. 2015; 197(5):983–91. doi: 10.1128/JB.02325-14 PMID: 25535275

36. Potrykus K, Cashel M. (p)ppGpp: still magical? Annu Rev Microbiol. 2008; 62:35–51. doi: 10.1146/annurev.micro.62.081307.162903 PMID: 18454629

37. Cashel M, Gallant J. Two compounds implicated in the function of the RC gene of Escherichia coli. Nature. 1969; 221(5183):838–41. PMID: 4885263

38. Barker MM, Gaal T, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. J Mol Biol. 2001; 305(4):689–702. PMID: 11162085

39. Paul BJ, Barker MM, Ross W, Schneider DA, Webb C, Foster JW, et al. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. Cell. 2004; 118(3):311–22. PMID: 15294157

40. Paul BJ, Berkmen MB, Gourse RL. DksA potentiates direct activation of amino acid promoters by ppGpp. Proc Natl Acad Sci U S A. 2005; 102(22):7823–8. PMID: 15899978

41. Mukherjee S, Yakhnin H, Kysela D, Sokoloski J, Babitzke P, Kearns DB. CsrA-FliW interaction governs flagellin homeostasis and a checkpoint on flagellar morphogenesis in Bacillus subtilis. Mol Microbiol. 2011; 82(2):447–61. doi: 10.1111/j.1365-2958.2011.07822.x PMID: 21895793

42. Uzzau S, Figueroa-Bossi N, Rubino S, Bossi L. Epitope tagging of chromosomal genes in Salmonella. Proc Natl Acad Sci U S A. 2001; 98(26):15264–9. PMID: 11742086

43. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000; 97(12):6640–5. PMID: 10829079

44. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24–6. doi: 10.1038/nbt.1754 PMID: 21221095

45. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14(2):178–92. doi: 10.1093/bib/bbs017 PMID: 22517427

46. Albert I, Wachi S, Jiang C, Pugh BF. GeneTrack—a genomic data processing and visualization framework. Bioinformatics. 2008; 24(10):1305–6. doi: 10.1093/bioinformatics/btn119 PMID: 18388141

47. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37(Web Server issue):W202–8. doi: 10.1093/nar/gkp335 PMID: 19458158

48. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS. 2008; 12(2):137–41. doi: 10.1089/omi.2008.0017 PMID: 18416670

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215(3):403–10. PMID: 2231712

50. Federhen S. The NCBI Taxonomy database. Nucleic Acids Res. 2012; 40(Database issue):D136–43. doi: 10.1093/nar/gkr1178 PMID: 22139910

51. Dias R, Xavier MG, Rossi FD, Neves MV, Lange TA, Giongo A, et al. MPI-blastn and NCBI-TaxCollector: improving metagenomic analysis with high performance classification and wide taxonomic attachment. J Bioinform Comput Biol. 2014; 12(3):1450013. doi: 10.1142/S0219720014500139 PMID: 24969751

52. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007; 23(1):127–8. PMID: 17050570

53. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014; 42(Database issue):D199–205. doi: 10.1093/nar/gkt1076 PMID: 24214961

54. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011; 2011:bar009.

55. Goosen N, van de Putte P. The regulation of transcription initiation by integration host factor. Mol Microbiol. 1995; 16(1):1–7. PMID: 7651128

56. Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell. 1996; 87(7):1295–306. PMID: 8980235

57. Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee AS, Luscombe NM. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in Escherichia coli K12. Nucleic Acids Res. 2012; 40(8):3524–37. doi: 10.1093/nar/gkr1236 PMID: 22180530

58. Humair B, Wackwitz B, Haas D. GacA-controlled activation of promoters for small RNA genes in Pseu-domonas fluorescens. Appl Environ Microbiol. 2010; 76(5):1497–506. doi: 10.1128/AEM.02014-09 PMID: 20048056

59. Schröder I, Darie S, Gunsalus RP. Activation of the Escherichia coli nitrate reductase (narGHJI) operon by NarL and Fnr requires integration host factor. J Biol Chem. 1993; 268(2):771–4. PMID: 8419352

60. Lee DJ, Minchin SD, Busby SJ. Activating transcription in bacteria. Annu Rev Microbiol. 2012; 66:125–52. doi: 10.1146/annurev-micro-092611-150012 PMID: 22726217

61. Richet E, Vidal-Ingigliardi D, Raibaud O. A new mechanism for coactivation of transcription initiation: repositioning of an activator triggered by the binding of a second activator. Cell. 1991; 66(6):1185–95. PMID: 1913806

62. Wade JT, Struhl K, Busby SJ, Grainger DC. Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. Mol Microbiol. 2007; 65(1):21–6. PMID: 17581117

63. Wade JT, Reppas NB, Church GM, Struhl K. Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. Genes Dev. 2005; 19 (21):2619–30. PMID: 16264194

64. Teif VB, Rippe K. Statistical-mechanical lattice models for protein-DNA binding in chromatin. J Phys Condens Matter. 2010; 22(41):414105. doi: 10.1088/0953-8984/22/41/414105 PMID: 21386588

65. Mrázek J, Xie S, Guo X, Srivastava A. AIMIE: a web-based environment for detection and interpretation of significant sequence motifs in prokaryotic genomes. Bioinformatics. 2008; 24(8):1041–8. doi: 10.1093/bioinformatics/btn077 PMID: 18304933

66. Vakulskas CA, Potts AH, Babitzke P, Ahmer BM, Romeo T. Regulation of bacterial virulence by Csr (Rsm) systems. Microbiol Mol Biol Rev. 2015; 79(2):193–224. doi: 10.1128/MMBR.00052-14 PMID: 25833324

67. Yakhnin H, Pandit P, Petty TJ, Baker CS, Romeo T, Babitzke P. CsrA of Bacillus subtilis regulates translation initiation of the gene encoding the flagellin protein (hag) by blocking ribosome binding. Mol Microbiol. 2007; 64(6):1605–20. PMID: 17555441

68. Dessimoz C, Gabaldón T, Roos DS, Sonnhammer EL, Herrero J, Consortium QfO. Toward community standards in the quest for orthologs. Bioinformatics. 2012; 28(6):900–4. doi: 10.1093/bioinformatics/bts050 PMID: 22332236